

DNA: Volume 1

Contents

Articles

Copyright@2009 by Bci2 1

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version ;published by the Free Software Foundation, with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in ;the section entitled "GNU Free Documentation License". Edited by Bci2, with all contributors listed after the License statement. 2

DNA Basics 3

DNA	3
DNA sequence	27
DNA sequencing	28
DNA profiling	37
DNA polymerase	49
DNA Topoisomerase	53
List of nucleic acid simulation software	56

DNA Structure and Functions 58

DNA structure	58
Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid	59
Molecular models of DNA	64
DNA replication	76
DNA repair	83
DNA Translation	95
DNA Transcription	100
DNA Transfer	106
Reverse transcriptase	109
DNA microarray	117
Triple-stranded DNA	126
G-quadruplex	127
DNA Analysis	131
Human genome	136

References

Article Sources and Contributors	144
Image Sources, Licenses and Contributors	147

Article Licenses

License	150
---------	-----

Copyright@2009 by Bci2

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version ;published by the Free Software Foundation, with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in ;the section entitled "GNU Free Documentation License". Edited by Bci2, with all contributors listed after the License statement.

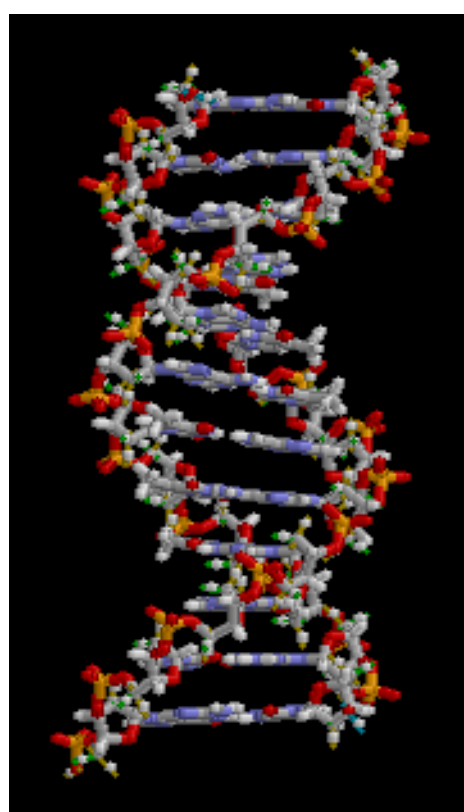
DNA Basics

DNA

Deoxyribonucleic acid ( /diˈɒksiːˈraɪboʊnuːkliːk ˈæsɪd/ Wikipedia:Media helpFile:en-us-Deoxyribonucleic_acid.ogg)

(**DNA**) is a nucleic acid that contains the genetic instructions used in the development and functioning of all known living organisms and some viruses. The main role of DNA molecules is the long-term storage of information. DNA is often compared to a set of blueprints or a recipe, or a code, since it contains the instructions needed to construct other components of cells, such as proteins and RNA molecules. The DNA segments that carry this genetic information are called genes, but other DNA sequences have structural purposes, or are involved in regulating the use of this genetic information.

Chemically, DNA consists of two long polymers of simple units called nucleotides, with backbones made of sugars and phosphate groups joined by ester bonds. These two strands run in opposite directions to each other and are therefore anti-parallel. Attached to each sugar is one of four types of molecules called bases. It is the sequence of these four bases along the backbone that encodes information. This information is read using the genetic code, which specifies the sequence of the amino acids within proteins. The code is read by copying stretches of DNA into the related nucleic acid RNA, in a process called transcription.



The structure of part of a DNA double helix

Within cells, DNA is organized into long structures called chromosomes. These chromosomes are duplicated before cells divide, in a process called DNA replication. Eukaryotic organisms (animals, plants, fungi, and protists) store most of their DNA inside the cell nucleus and some of their DNA in organelles, such as mitochondria or chloroplasts.^[1] In contrast, prokaryotes (bacteria and archaea) store their DNA only in the cytoplasm. Within the chromosomes, chromatin proteins such as histones compact and organize DNA. These compact structures guide the interactions between DNA and other proteins, helping control which parts of the DNA are transcribed.

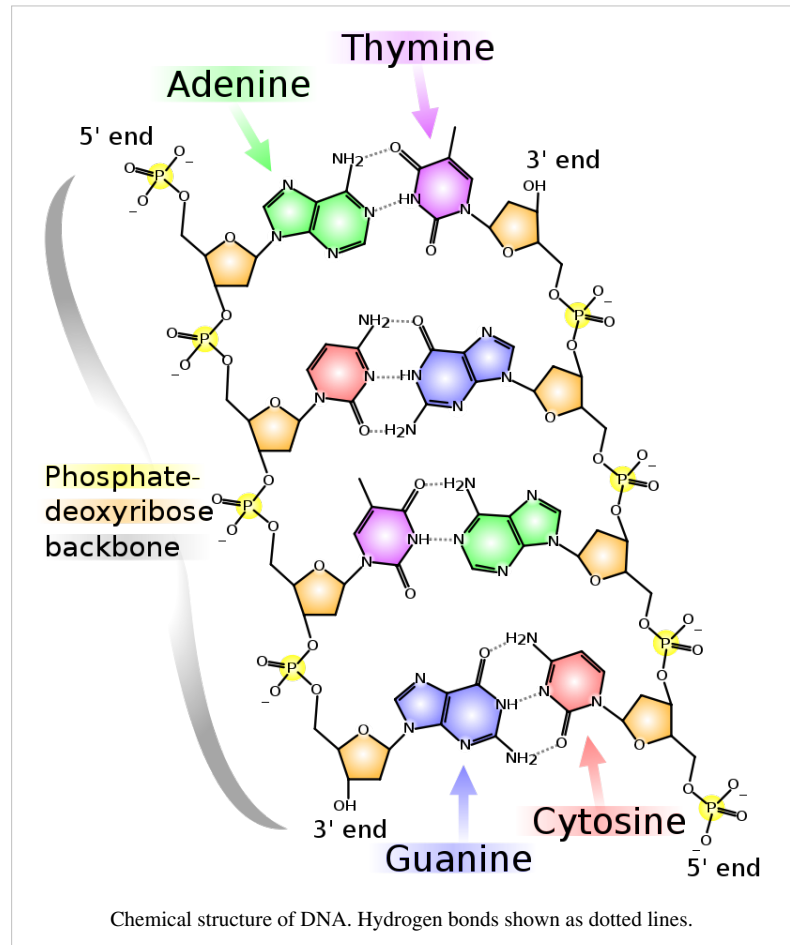
Properties

DNA is a long polymer made from repeating units called nucleotides.^{[2] [3] [4]} The DNA chain is 22 to 26 Ångströms wide (2.2 to 2.6 nanometres), and one nucleotide unit is 3.3 Å (0.33 nm) long.^[5] Although each individual repeating unit is very small, DNA polymers can be very large molecules containing millions of nucleotides. For instance, the largest human chromosome, chromosome number 1, is approximately 220 million base pairs long.^[6]

In living organisms, DNA does not usually exist as a single molecule, but instead as a pair of molecules that are held tightly together.^{[7] [8]} These two long strands entwine like vines, in the shape of a double helix. The nucleotide repeats contain both the segment of the backbone of the molecule, which holds the chain together, and a base, which interacts with the other DNA strand in the helix. A base linked to a sugar is

called a nucleoside and a base linked to a sugar and one or more phosphate groups is called a nucleotide. If multiple nucleotides are linked together, as in DNA, this polymer is called a polynucleotide.^[9]

The backbone of the DNA strand is made from alternating phosphate and sugar residues.^[10] The sugar in DNA is 2-deoxyribose, which is a pentose (five-carbon) sugar. The sugars are joined together by phosphate groups that form phosphodiester bonds between the third and fifth carbon atoms of adjacent sugar rings. These asymmetric bonds mean a strand of DNA has a direction. In a double helix the direction of the nucleotides in one strand is opposite to their direction in the other strand: the strands are *antiparallel*. The asymmetric ends of DNA strands are called the 5' (*five prime*) and 3' (*three prime*) ends, with the 5' end having a terminal phosphate group and the 3' end a terminal hydroxyl group. One major difference between DNA and RNA is the sugar, with the 2-deoxyribose in DNA being replaced by the alternative pentose sugar ribose in RNA.^[8]



The DNA double helix is stabilized by hydrogen bonds between the bases attached to the two strands. The four bases found in DNA are adenine (abbreviated A), cytosine (C), guanine (G) and thymine (T). These four bases are attached to the sugar/phosphate to form the complete nucleotide, as shown for adenosine monophosphate.

These bases are classified into two types; adenine and guanine are fused five- and six-membered heterocyclic compounds called purines, while cytosine and thymine are six-membered rings called pyrimidines.^[8] A fifth pyrimidine base, called uracil (U), usually takes the place of thymine in RNA and differs from thymine by lacking a methyl group on its ring. Uracil is not usually found in DNA, occurring only as a breakdown product of cytosine. In addition to RNA and DNA, a large number of artificial nucleic acid analogues have also been created to study the properties of nucleic acids, or for use in biotechnology.^[12]

Grooves

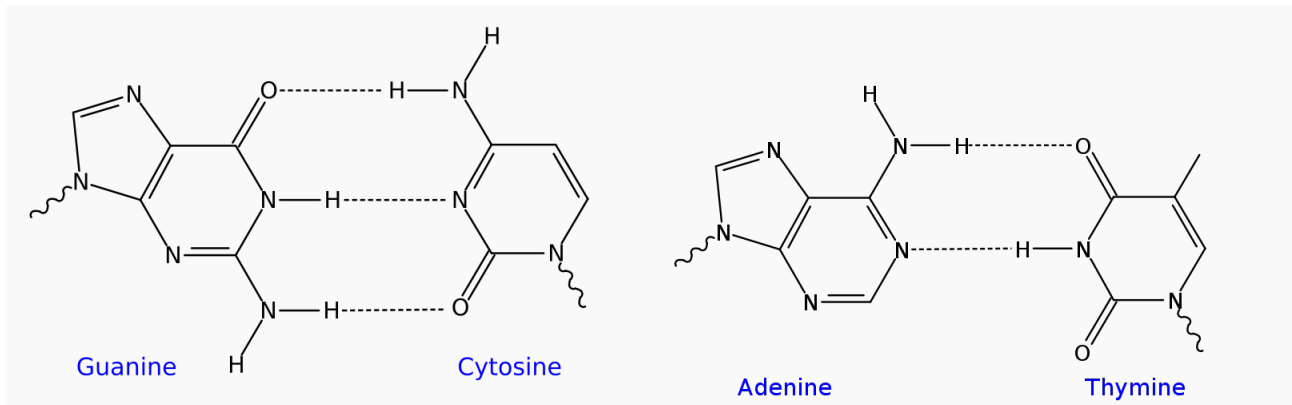
Twin helical strands form the DNA backbone. Another double helix may be found by tracing the spaces, or grooves, between the strands. These voids are adjacent to the base pairs and may provide a binding site. As the strands are not directly opposite each other, the grooves are unequally sized. One groove, the major groove, is 22 Å wide and the other, the minor groove, is 12 Å wide.^[13] The narrowness of the minor groove means that the edges of the bases are more accessible in the major groove. As a result, proteins like transcription factors that can bind to specific sequences in double-stranded DNA usually make contacts to the sides of the bases exposed in the major groove.^[14] This situation varies in unusual conformations of DNA within the cell (*see below*), but the major and minor grooves are always named to reflect the differences in size that would be seen if the DNA is twisted back into the ordinary B form.

Base pairing

Each type of base on one strand forms a bond with just one type of base on the other strand. This is called complementary base pairing. Here, purines form hydrogen bonds to pyrimidines, with A bonding only to T, and C bonding only to G. This arrangement of two nucleotides binding together across the double helix is called a base pair. As hydrogen bonds are not covalent, they can be broken and rejoined relatively easily. The two strands of DNA in a double helix can therefore be pulled apart like a zipper, either by a mechanical force or high temperature.^[15] As a result of this complementarity, all the information in the double-stranded sequence of a DNA helix is duplicated on each strand, which is vital in DNA replication. Indeed, this reversible and specific interaction between complementary base pairs is critical for all the functions of DNA in living organisms.^[3]



A section of DNA. The bases lie horizontally between the two spiraling strands.^[11] Animated version at File:DNA orbit animated.gif.



Top, a **GC** base pair with three hydrogen bonds. Bottom, an **AT** base pair with two hydrogen bonds. Non-covalent hydrogen bonds between the pairs are shown as dashed lines.

The two types of base pairs form different numbers of hydrogen bonds, AT forming two hydrogen bonds, and GC forming three hydrogen bonds (see figures, left). DNA with high GC-content is more stable than DNA with low GC-content, but contrary to popular belief, this is not due to the extra hydrogen bond of a GC base pair but rather the contribution of stacking interactions (hydrogen bonding merely provides specificity of the pairing, not stability).^[16] As a result, it is both the percentage of GC base pairs and the overall length of a DNA double helix that determine the strength of the association between the two strands of DNA. Long DNA helices with a high GC content have stronger-interacting strands, while short helices with high AT content have weaker-interacting strands.^[17] In biology, parts of the DNA double helix that need to separate easily, such as the TATAAT Pribnow box in some promoters, tend to have a high AT content, making the strands easier to pull apart.^[18] In the laboratory, the strength of this interaction can be measured by finding the temperature required to break the hydrogen bonds, their melting temperature (also called T_m value). When all the base pairs in a DNA double helix melt, the strands separate and exist in solution as two entirely independent molecules. These single-stranded DNA molecules have no single common shape, but some conformations are more stable than others.^[19]

Sense and antisense

A DNA sequence is called "sense" if its sequence is the same as that of a messenger RNA copy that is translated into protein.^[20] The sequence on the opposite strand is called the "antisense" sequence. Both sense and antisense sequences can exist on different parts of the same strand of DNA (i.e. both strands contain both sense and antisense sequences). In both prokaryotes and eukaryotes, antisense RNA sequences are produced, but the functions of these RNAs are not entirely clear.^[21] One proposal is that antisense RNAs are involved in regulating gene expression through RNA-RNA base pairing.^[22]

A few DNA sequences in prokaryotes and eukaryotes, and more in plasmids and viruses, blur the distinction between sense and antisense strands by having overlapping genes.^[23] In these cases, some DNA sequences do double duty, encoding one protein when read along one strand, and a second protein when read in the opposite direction along the other strand. In bacteria, this overlap may be involved in the regulation of gene transcription,^[24] while in viruses, overlapping genes increase the amount of information that can be encoded within the small viral genome.^[25]

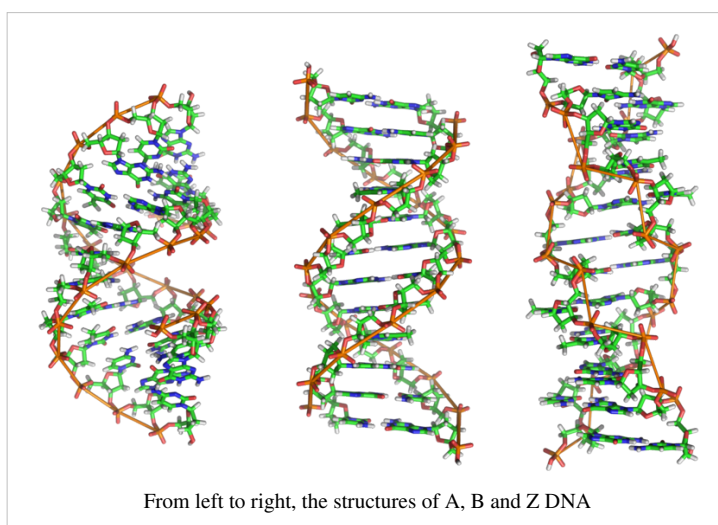
Supercoiling

DNA can be twisted like a rope in a process called DNA supercoiling. With DNA in its "relaxed" state, a strand usually circles the axis of the double helix once every 10.4 base pairs, but if the DNA is twisted the strands become more tightly or more loosely wound.^[26] If the DNA is twisted in the direction of the helix, this is positive supercoiling, and the bases are held more tightly together. If they are twisted in the opposite direction, this is negative supercoiling, and the bases come apart more easily. In nature, most DNA has slight negative supercoiling that is introduced by enzymes called topoisomerases.^[27] These enzymes are also needed to relieve the twisting

stresses introduced into DNA strands during processes such as transcription and DNA replication.^[28]

Alternate DNA structures

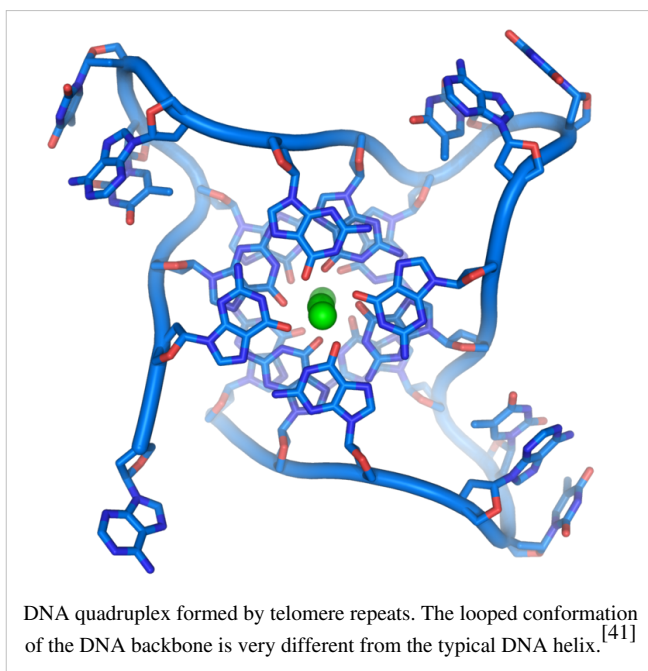
DNA exists in many possible conformations that include A-DNA, B-DNA, and Z-DNA forms, although, only B-DNA and Z-DNA have been directly observed in functional organisms.^[10] The conformation that DNA adopts depends on the hydration level, DNA sequence, the amount and direction of supercoiling, chemical modifications of the bases, the type and concentration of metal ions, as well as the presence of polyamines in solution.^[29]



The first published reports of A-DNA X-ray diffraction patterns— and also B-DNA used analyses based on Patterson transforms that provided only a limited amount of structural information for oriented fibers of DNA.^{[30] [31]} An alternate analysis was then proposed by Wilkins *et al.*, in 1953, for the *in vivo* B-DNA X-ray diffraction/scattering patterns of highly hydrated DNA fibers in terms of squares of Bessel functions.^[32] In the same journal, Watson and Crick presented their molecular modeling analysis of the DNA X-ray diffraction patterns to suggest that the structure was a double-helix.^[7]

Although the 'B-DNA form' is most common under the conditions found in cells,^[33] it is not a well-defined conformation but a family of related DNA conformations^[34] that occur at the high hydration levels present in living cells. Their corresponding X-ray diffraction and scattering patterns are characteristic of molecular paracrystals with a significant degree of disorder.^{[35] [36]}

Compared to B-DNA, the A-DNA form is a wider right-handed spiral, with a shallow, wide minor groove and a narrower, deeper major groove. The A form occurs under non-physiological conditions in partially dehydrated samples of DNA, while in the cell it may be produced in hybrid pairings of DNA and RNA strands, as well as in enzyme-DNA complexes.^{[37] [38]} Segments of DNA where the bases have been chemically modified by methylation may undergo a larger change in conformation and adopt the Z form. Here, the strands turn about the helical axis in a left-handed spiral, the opposite of the more common B form.^[39] These unusual structures can be recognized by specific Z-DNA binding proteins and may be involved in the regulation of transcription.^[40]



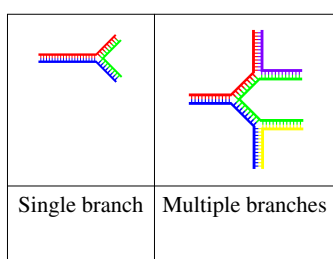
Quadruplex structures

At the ends of the linear chromosomes are specialized regions of DNA called telomeres. The main function of these regions is to allow the cell to replicate chromosome ends using the enzyme telomerase, as the enzymes that normally replicate DNA cannot copy the extreme 3' ends of chromosomes.^[42] These specialized chromosome caps also help protect the DNA ends, and stop the DNA repair systems in the cell from treating them as damage to be corrected.^[43] In human cells, telomeres are usually lengths of single-stranded DNA containing several thousand repeats of a simple TTAGGG sequence.^[44]

These guanine-rich sequences may stabilize chromosome ends by forming structures of stacked sets of four-base units, rather than the usual base

pairs found in other DNA molecules. Here, four guanine bases form a flat plate and these flat four-base units then stack on top of each other, to form a stable *G-quadruplex* structure.^[45] These structures are stabilized by hydrogen bonding between the edges of the bases and chelation of a metal ion in the centre of each four-base unit.^[46] Other structures can also be formed, with the central set of four bases coming from either a single strand folded around the bases, or several different parallel strands, each contributing one base to the central structure.

In addition to these stacked structures, telomeres also form large loop structures called telomere loops, or T-loops. Here, the single-stranded DNA curls around in a long circle stabilized by telomere-binding proteins.^[47] At the very end of the T-loop, the single-stranded telomere DNA is held onto a region of double-stranded DNA by the telomere strand disrupting the double-helical DNA and base pairing to one of the two strands. This triple-stranded structure is called a displacement loop or D-loop.^[45]

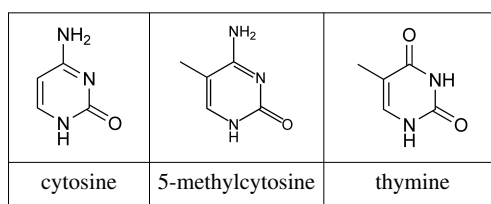


Branched DNA can form networks containing multiple branches.

Branched DNA

In DNA fraying occurs when non-complementary regions exist at the end of an otherwise complementary double-strand of DNA. However, branched DNA can occur if a third strand of DNA is introduced and contains adjoining regions able to hybridize with the frayed regions of the pre-existing double-strand. Although the simplest example of branched DNA involves only three strands of DNA, complexes involving additional strands and multiple branches are also possible.^[48] Branched DNA can be used in nanotechnology to construct geometric shapes, see the section on uses in technology below.

Chemical modifications



Structure of cytosine with and without the 5-methyl group. Deamination converts 5-methylcytosine into thymine.

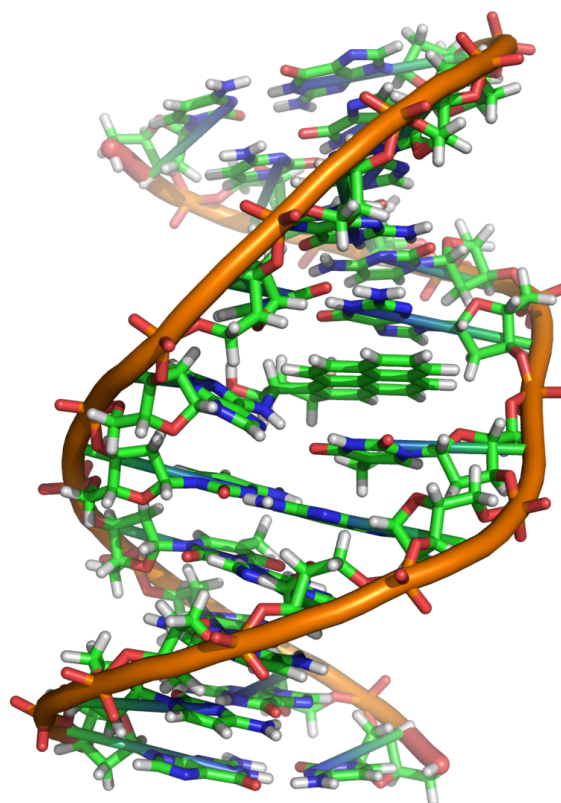
Base modifications

The expression of genes is influenced by how the DNA is packaged in chromosomes, in a structure called chromatin. Base modifications can be involved in packaging, with regions that have low or no gene expression usually containing high levels of methylation of cytosine bases. For example, cytosine methylation, produces 5-methylcytosine, which is important for X-chromosome inactivation.^[49] The average level of methylation varies between organisms - the worm *Caenorhabditis elegans* lacks cytosine methylation, while vertebrates have higher levels, with up to 1% of their DNA containing 5-methylcytosine.^[50] Despite the importance of 5-methylcytosine, it can deaminate to leave a thymine base, methylated cytosines are therefore particularly prone to mutations.^[51] Other base modifications include adenine methylation in bacteria, the presence of 5-hydroxymethylcytosine in the brain,^[52] and the glycosylation of uracil to produce the "J-base" in kinetoplasts.^[53] ^[54]

Damage

DNA can be damaged by many sorts of mutagens, which change the DNA sequence. Mutagens include oxidizing agents, alkylating agents and also high-energy electromagnetic radiation such as ultraviolet light and X-rays. The type of DNA damage produced depends on the type of mutagen. For example, UV light can damage DNA by producing thymine dimers, which are cross-links between pyrimidine bases.^[56] On the other hand, oxidants such as free radicals or hydrogen peroxide produce multiple forms of damage, including base modifications, particularly of guanosine, and double-strand breaks.^[57] A typical human cell contains about 150,000 bases that have suffered oxidative damage.^[58] Of these oxidative lesions, the most dangerous are double-strand breaks, as these are difficult to repair and can produce point mutations, insertions and deletions from the DNA sequence, as well as chromosomal translocations.^[59]

Many mutagens fit into the space between two adjacent base pairs, this is called *intercalating*. Most intercalators are aromatic and planar molecules, and include Ethidium bromide, daunomycin, and



A covalent adduct between benzo[*a*]pyrene, the major mutagen in tobacco smoke, and DNA^[55]

doxorubicin. In order for an intercalator to fit between base pairs, the bases must separate, distorting the DNA strands by unwinding of the double helix. This inhibits both transcription and DNA replication, causing toxicity and mutations. As a result, DNA intercalators are often carcinogens, and Benzo[*a*]pyrene diol epoxide, acridines, aflatoxin and ethidium bromide are well-known examples.^{[60] [61] [62]} Nevertheless, due to their ability to inhibit DNA transcription and replication, other similar toxins are also used in chemotherapy to inhibit rapidly growing cancer cells.^[63]

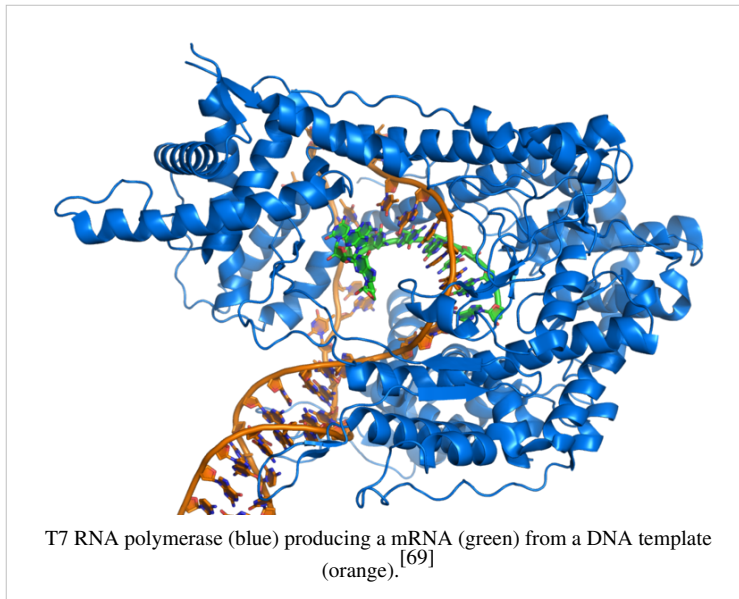
Biological functions

DNA usually occurs as linear chromosomes in eukaryotes, and circular chromosomes in prokaryotes. The set of chromosomes in a cell makes up its genome; the human genome has approximately 3 billion base pairs of DNA arranged into 46 chromosomes.^[64] The information carried by DNA is held in the sequence of pieces of DNA called genes. Transmission of genetic information in genes is achieved via complementary base pairing. For example, in transcription, when a cell uses the information in a gene, the DNA sequence is copied into a complementary RNA sequence through the attraction between the DNA and the correct RNA nucleotides. Usually, this RNA copy is then used to make a matching protein sequence in a process called translation which depends on the same interaction between RNA nucleotides. Alternatively, a cell may simply copy its genetic information in a process called DNA replication. The details of these functions are covered in other articles; here we focus on the interactions between DNA and other molecules that mediate the function of the genome.

Genes and genomes

Genomic DNA is located in the cell nucleus of eukaryotes, as well as small amounts in mitochondria and chloroplasts. In prokaryotes, the DNA is held within an irregularly shaped body in the cytoplasm called the nucleoid.^[65] The genetic information in a genome is held within genes, and the complete set of this information in an organism is called its genotype. A gene is a unit of heredity and is a region of DNA that influences a particular characteristic in an organism. Genes contain an open reading frame that can be transcribed, as well as regulatory sequences such as promoters and enhancers, which control the transcription of the open reading frame.

In many species, only a small fraction of the total sequence of the genome encodes protein. For example, only about 1.5% of the human genome consists of protein-coding exons, with over 50% of human DNA consisting of non-coding repetitive sequences.^[66] The reasons for the presence of so much non-coding DNA in eukaryotic genomes and the extraordinary differences in genome size, or *C-value*, among species represent a long-standing puzzle known as the "C-value enigma."^[67] However, DNA sequences that do not code protein may still encode functional non-coding RNA molecules, which are involved in the regulation of gene expression.^[68]



Some non-coding DNA sequences play structural roles in chromosomes. Telomeres and centromeres typically contain few genes, but are important for the function and stability of chromosomes.^{[43] [70]} An abundant form of non-coding DNA in humans are pseudogenes, which are copies of genes that have been disabled by mutation.^[71] These sequences are usually just molecular fossils, although they can occasionally serve as raw genetic material for the creation of new genes through the process of gene duplication and divergence.^[72]

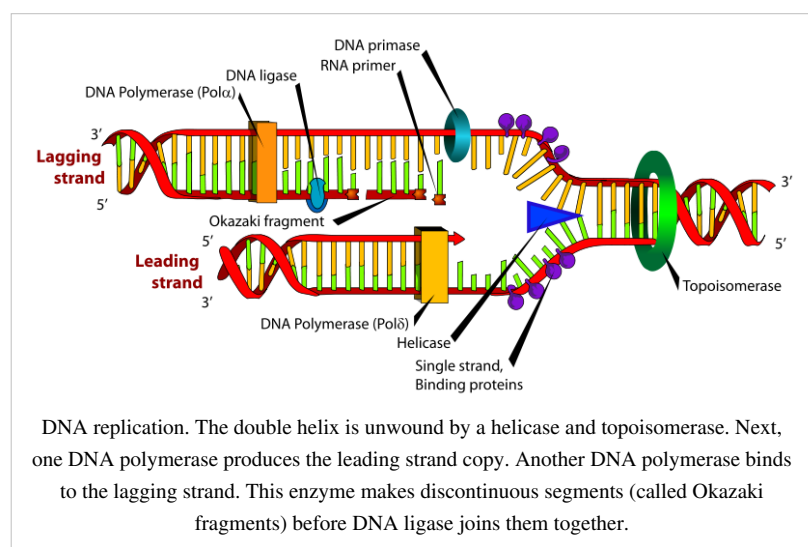
Transcription and translation

A gene is a sequence of DNA that contains genetic information and can influence the phenotype of an organism. Within a gene, the sequence of bases along a DNA strand defines a messenger RNA sequence, which then defines one or more protein sequences. The relationship between the nucleotide sequences of genes and the amino-acid sequences of proteins is determined by the rules of translation, known collectively as the genetic code. The genetic code consists of three-letter 'words' called *codons* formed from a sequence of three nucleotides (e.g. ACT, CAG, TTT).

In transcription, the codons of a gene are copied into messenger RNA by RNA polymerase. This RNA copy is then decoded by a ribosome that reads the RNA sequence by base-pairing the messenger RNA to transfer RNA, which carries amino acids. Since there are 4 bases in 3-letter combinations, there are 64 possible codons (4^3 combinations). These encode the twenty standard amino acids, giving most amino acids more than one possible codon. There are also three 'stop' or 'nonsense' codons signifying the end of the coding region; these are the TAA, TGA and TAG codons.

Replication

Cell division is essential for an organism to grow, but when a cell divides it must replicate the DNA in its genome so that the two daughter cells have the same genetic information as their parent. The double-stranded structure of DNA provides a simple mechanism for DNA replication. Here, the two strands are separated and then each strand's complementary DNA sequence is recreated by an enzyme called DNA polymerase. This enzyme makes the complementary strand by



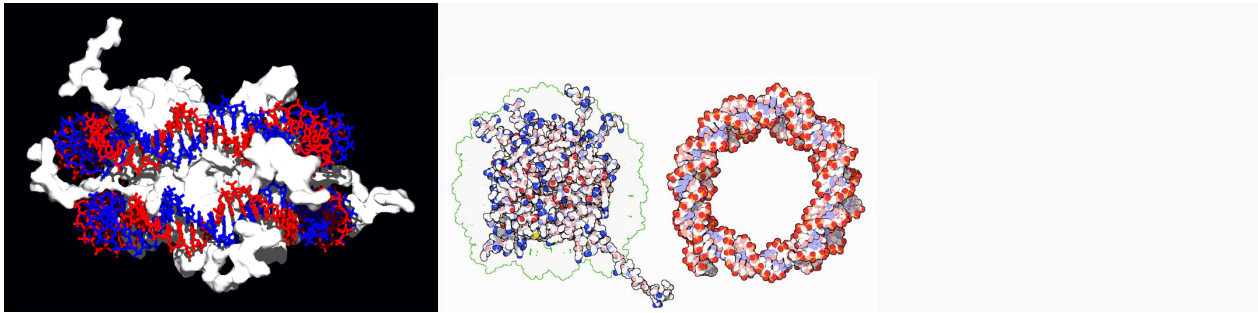
finding the correct base through complementary base pairing, and bonding it onto the original strand. As DNA polymerases can only extend a DNA strand in a 5' to 3' direction, different mechanisms are used to copy the

antiparallel strands of the double helix.^[73] In this way, the base on the old strand dictates which base appears on the new strand, and the cell ends up with a perfect copy of its DNA.

Interactions with proteins

All the functions of DNA depend on interactions with proteins. These protein interactions can be non-specific, or the protein can bind specifically to a single DNA sequence. Enzymes can also bind to DNA and of these, the polymerases that copy the DNA base sequence in transcription and DNA replication are particularly important.

DNA-binding proteins



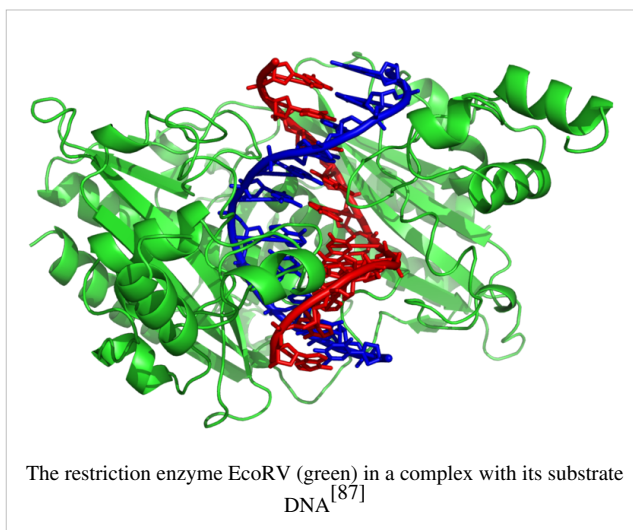
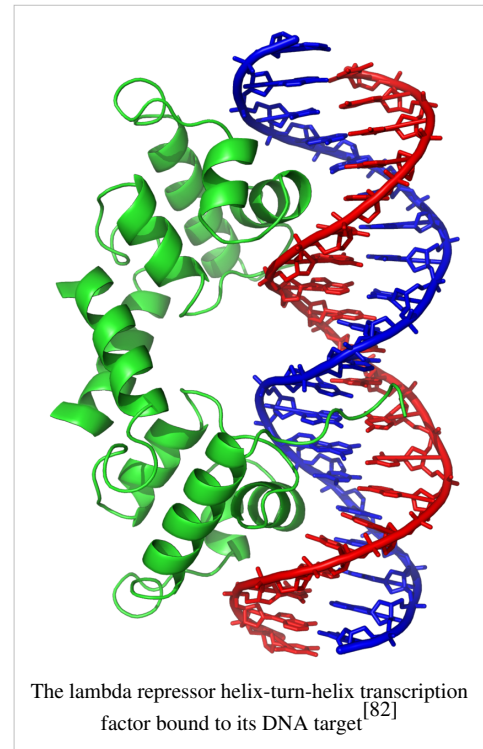
Interaction of DNA with histones (shown in white, top). These proteins' basic amino acids (below left, blue) bind to the acidic phosphate groups on DNA (below right, red).

Structural proteins that bind DNA are well-understood examples of non-specific DNA-protein interactions. Within chromosomes, DNA is held in complexes with structural proteins. These proteins organize the DNA into a compact structure called chromatin. In eukaryotes this structure involves DNA binding to a complex of small basic proteins called histones, while in prokaryotes multiple types of proteins are involved.^{[74] [75]} The histones form a disk-shaped complex called a nucleosome, which contains two complete turns of double-stranded DNA wrapped around its surface. These non-specific interactions are formed through basic residues in the histones making ionic bonds to the acidic sugar-phosphate backbone of the DNA, and are therefore largely independent of the base sequence.^[76] Chemical modifications of these basic amino acid residues include methylation, phosphorylation and acetylation.^[77] These chemical changes alter the strength of the interaction between the DNA and the histones, making the DNA more or less accessible to transcription factors and changing the rate of transcription.^[78] Other non-specific DNA-binding proteins in chromatin include the high-mobility group proteins, which bind to bent or distorted DNA.^[79] These proteins are important in bending arrays of nucleosomes and arranging them into the larger structures that make up chromosomes.^[80]

A distinct group of DNA-binding proteins are the DNA-binding proteins that specifically bind single-stranded DNA. In humans, replication protein A is the best-understood member of this family and is used in processes where the double helix is separated, including DNA replication, recombination and DNA repair.^[81] These binding proteins seem to stabilize single-stranded DNA and protect it from forming stem-loops or being degraded by nucleases.

In contrast, other proteins have evolved to bind to particular DNA sequences. The most intensively studied of these are the various transcription factors, which are proteins that regulate transcription. Each transcription factor binds to one particular set of DNA sequences and activates or inhibits the transcription of genes that have these sequences close to their promoters. The transcription factors do this in two ways. Firstly, they can bind the RNA polymerase responsible for transcription, either directly or through other mediator proteins; this locates the polymerase at the promoter and allows it to begin transcription.^[83] Alternatively, transcription factors can bind enzymes that modify the histones at the promoter; this will change the accessibility of the DNA template to the polymerase.^[84]

As these DNA targets can occur throughout an organism's genome, changes in the activity of one type of transcription factor can affect thousands of genes.^[85] Consequently, these proteins are often the targets of the signal transduction processes that control responses to environmental changes or cellular differentiation and development. The specificity of these transcription factors' interactions with DNA come from the proteins making multiple contacts to the edges of the DNA bases, allowing them to "read" the DNA sequence. Most of these base-interactions are made in the major groove, where the bases are most accessible.^[86]



DNA-modifying enzymes

Nucleases and ligases

Nucleases are enzymes that cut DNA strands by catalyzing the hydrolysis of the phosphodiester bonds. Nucleases that hydrolyse nucleotides from the ends of DNA strands are called exonucleases, while endonucleases cut within strands. The most frequently used nucleases in molecular biology are the restriction endonucleases, which cut DNA at specific sequences. For instance, the EcoRV enzyme shown to the left recognizes the 6-base sequence 5'-GAT|ATC-3' and makes a cut at the vertical line. In nature, these enzymes protect bacteria against phage infection by

digesting the phage DNA when it enters the bacterial cell, acting as part of the restriction modification system.^[88] In technology, these sequence-specific nucleases are used in molecular cloning and DNA fingerprinting.

Enzymes called DNA ligases can rejoin cut or broken DNA strands.^[89] Ligases are particularly important in lagging strand DNA replication, as they join together the short segments of DNA produced at the replication fork into a complete copy of the DNA template. They are also used in DNA repair and genetic recombination.^[89]

Topoisomerases and helicases

Topoisomerases are enzymes with both nuclease and ligase activity. These proteins change the amount of supercoiling in DNA. Some of these enzymes work by cutting the DNA helix and allowing one section to rotate, thereby reducing its level of supercoiling; the enzyme then seals the DNA break.^[27] Other types of these enzymes are capable of cutting one DNA helix and then passing a second strand of DNA through this break, before rejoining the helix.^[90] Topoisomerases are required for many processes involving DNA, such as DNA replication and transcription.^[28]

Helicases are proteins that are a type of molecular motor. They use the chemical energy in nucleoside triphosphates, predominantly ATP, to break hydrogen bonds between bases and unwind the DNA double helix into single strands.^[91] These enzymes are essential for most processes where enzymes need to access the DNA bases.

Polymerases

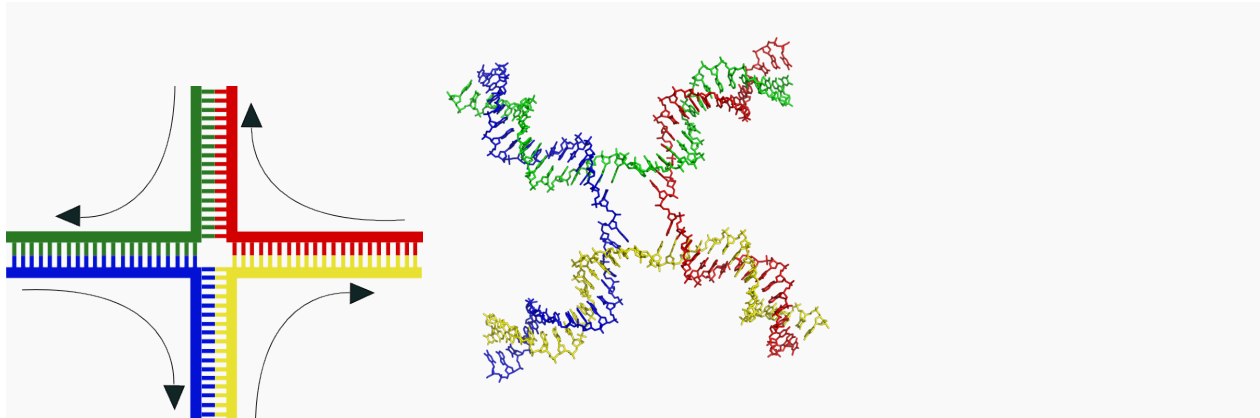
Polymerases are enzymes that synthesize polynucleotide chains from nucleoside triphosphates. The sequence of their products are copies of existing polynucleotide chains - which are called *templates*. These enzymes function by adding nucleotides onto the 3' hydroxyl group of the previous nucleotide in a DNA strand. Consequently, all polymerases work in a 5' to 3' direction.^[92] In the active site of these enzymes, the incoming nucleoside triphosphate base-pairs to the template: this allows polymerases to accurately synthesize the complementary strand of their template. Polymerases are classified according to the type of template that they use.

In DNA replication, a DNA-dependent DNA polymerase makes a copy of a DNA sequence. Accuracy is vital in this process, so many of these polymerases have a proofreading activity. Here, the polymerase recognizes the occasional mistakes in the synthesis reaction by the lack of base pairing between the mismatched nucleotides. If a mismatch is detected, a 3' to 5' exonuclease activity is activated and the incorrect base removed.^[93] In most organisms DNA polymerases function in a large complex called the replisome that contains multiple accessory subunits, such as the DNA clamp or helicases.^[94]

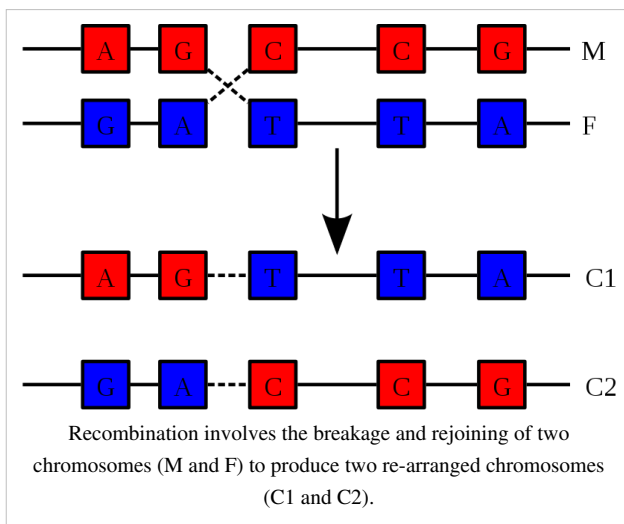
RNA-dependent DNA polymerases are a specialized class of polymerases that copy the sequence of an RNA strand into DNA. They include reverse transcriptase, which is a viral enzyme involved in the infection of cells by retroviruses, and telomerase, which is required for the replication of telomeres.^[42] ^[95] Telomerase is an unusual polymerase because it contains its own RNA template as part of its structure.^[43]

Transcription is carried out by a DNA-dependent RNA polymerase that copies the sequence of a DNA strand into RNA. To begin transcribing a gene, the RNA polymerase binds to a sequence of DNA called a promoter and separates the DNA strands. It then copies the gene sequence into a messenger RNA transcript until it reaches a region of DNA called the terminator, where it halts and detaches from the DNA. As with human DNA-dependent DNA polymerases, RNA polymerase II, the enzyme that transcribes most of the genes in the human genome, operates as part of a large protein complex with multiple regulatory and accessory subunits.^[96]

Genetic recombination



Structure of the Holliday junction intermediate in genetic recombination. The four separate DNA strands are coloured red, blue, green and yellow.^[97]



A DNA helix usually does not interact with other segments of DNA, and in human cells the different chromosomes even occupy separate areas in the nucleus called "chromosome territories".^[98] This physical separation of different chromosomes is important for the ability of DNA to function as a stable repository for information, as one of the few times chromosomes interact is during chromosomal crossover when they recombine. Chromosomal crossover is when two DNA helices break, swap a section and then rejoin.

Recombination allows chromosomes to exchange genetic information and produces new combinations of genes, which increases the efficiency of natural selection and can be important in the rapid evolution of

new proteins.^[99] Genetic recombination can also be involved in DNA repair, particularly in the cell's response to double-strand breaks.^[100]

The most common form of chromosomal crossover is homologous recombination, where the two chromosomes involved share very similar sequences. Non-homologous recombination can be damaging to cells, as it can produce chromosomal translocations and genetic abnormalities. The recombination reaction is catalyzed by enzymes known as recombinases, such as RAD51.^[101] The first step in recombination is a double-stranded break either caused by an endonuclease or damage to the DNA.^[102] A series of steps catalyzed in part by the recombinase then leads to joining of the two helices by at least one Holliday junction, in which a segment of a single strand in each helix is annealed to the complementary strand in the other helix. The Holliday junction is a tetrahedral junction structure that can be moved along the pair of chromosomes, swapping one strand for another. The recombination reaction is then halted by cleavage of the junction and re-ligation of the released DNA.^[103]

Evolution

DNA contains the genetic information that allows all modern living things to function, grow and reproduce. However, it is unclear how long in the 4-billion-year history of life DNA has performed this function, as it has been proposed that the earliest forms of life may have used RNA as their genetic material.^{[92] [104]} RNA may have acted as the central part of early cell metabolism as it can both transmit genetic information and carry out catalysis as part of ribozymes.^[105] This ancient RNA world where nucleic acid would have been used for both catalysis and genetics may have influenced the evolution of the current genetic code based on four nucleotide bases. This would occur since the number of different bases in such an organism is a trade-off between a small number of bases increasing replication accuracy and a large number of bases increasing the catalytic efficiency of ribozymes.^[106]

Unfortunately, there is no direct evidence of ancient genetic systems, as recovery of DNA from most fossils is impossible. This is because DNA will survive in the environment for less than one million years and slowly degrades into short fragments in solution.^[107] Claims for older DNA have been made, most notably a report of the isolation of a viable bacterium from a salt crystal 250 million years old,^[108] but these claims are controversial.^{[109] [110]}

Uses in technology

Genetic engineering

Methods have been developed to purify DNA from organisms, such as phenol-chloroform extraction and manipulate it in the laboratory, such as restriction digests and the polymerase chain reaction. Modern biology and biochemistry make intensive use of these techniques in recombinant DNA technology. Recombinant DNA is a man-made DNA sequence that has been assembled from other DNA sequences. They can be transformed into organisms in the form of plasmids or in the appropriate format, by using a viral vector.^[111] The genetically modified organisms produced can be used to produce products such as recombinant proteins, used in medical research,^[112] or be grown in agriculture.^{[113] [114]}

Forensics

Forensic scientists can use DNA in blood, semen, skin, saliva or hair found at a crime scene to identify a matching DNA of an individual, such as a perpetrator. This process is called genetic fingerprinting, or more accurately, DNA profiling. In DNA profiling, the lengths of variable sections of repetitive DNA, such as short tandem repeats and minisatellites, are compared between people. This method is usually an extremely reliable technique for identifying a matching DNA.^[115] However, identification can be complicated if the scene is contaminated with DNA from several people.^[116] DNA profiling was developed in 1984 by British geneticist Sir Alec Jeffreys,^[117] and first used in forensic science to convict Colin Pitchfork in the 1988 Enderby murders case.^[118]

People convicted of certain types of crimes may be required to provide a sample of DNA for a database. This has helped investigators solve old cases where only a DNA sample was obtained from the scene. DNA profiling can also be used to identify victims of mass casualty incidents.^[119] On the other hand, many convicted people have been released from prison on the basis of DNA techniques, which were not available when a crime had originally been committed.

Bioinformatics

Bioinformatics involves the manipulation, searching, and data mining of DNA sequence data. The development of techniques to store and search DNA sequences have led to widely applied advances in computer science, especially string searching algorithms, machine learning and database theory.^[120] String searching or matching algorithms, which find an occurrence of a sequence of letters inside a larger sequence of letters, were developed to search for specific sequences of nucleotides.^[121] In other applications such as text editors, even simple algorithms for this

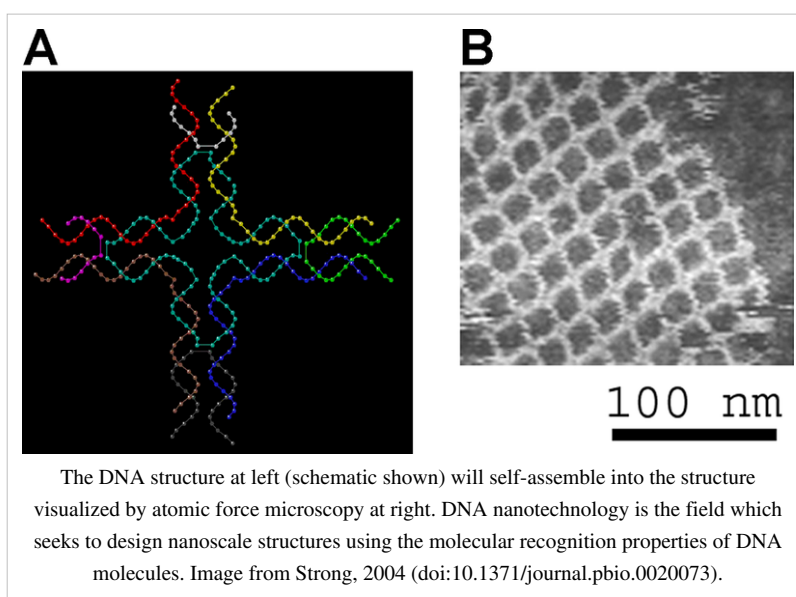
problem usually suffice, but DNA sequences cause these algorithms to exhibit near-worst-case behaviour due to their small number of distinct characters. The related problem of sequence alignment aims to identify homologous sequences and locate the specific mutations that make them distinct. These techniques, especially multiple sequence alignment, are used in studying phylogenetic relationships and protein function.^[122] Data sets representing entire genomes' worth of DNA sequences, such as those produced by the Human Genome Project, are difficult to use without annotations, which label the locations of genes and regulatory elements on each chromosome. Regions of DNA sequence that have the characteristic patterns associated with protein- or RNA-coding genes can be identified by gene finding algorithms, which allow researchers to predict the presence of particular gene products in an organism even before they have been isolated experimentally.^[123]

DNA nanotechnology

DNA nanotechnology uses the unique molecular recognition properties of DNA and other nucleic acids to create self-assembling branched DNA complexes with useful properties.^[124]

DNA is thus used as a structural material rather than as a carrier of biological information. This has led to the creation of two-dimensional periodic lattices (both tile-based as well as using the "DNA origami" method) as well as three-dimensional structures in the shapes of polyhedra.^[125] Nanomechanical devices and algorithmic self-assembly have also been demonstrated,^[126] and

these DNA structures have been used to template the arrangement of other molecules such as gold nanoparticles and streptavidin proteins.^[127]



History and anthropology

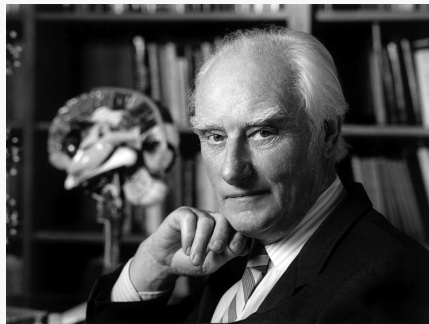
Because DNA collects mutations over time, which are then inherited, it contains historical information and by comparing DNA sequences, geneticists can infer the evolutionary history of organisms, their phylogeny.^[128] This field of phylogenetics is a powerful tool in evolutionary biology. If DNA sequences within a species are compared, population geneticists can learn the history of particular populations. This can be used in studies ranging from ecological genetics to anthropology; for example, DNA evidence is being used to try to identify the Ten Lost Tribes of Israel.^{[129] [130]}

DNA has also been used to look at modern family relationships, such as establishing family relationships between the descendants of Sally Hemings and Thomas Jefferson. This usage is closely related to the use of DNA in criminal investigations detailed above. Indeed, some criminal investigations have been solved when DNA from crime scenes has matched relatives of the guilty individual.^[131]

History of DNA research

DNA was first isolated by the Swiss physician Friedrich Miescher who, in 1869, discovered a microscopic substance in the pus of discarded surgical bandages. As it resided in the nuclei of cells, he called it "nuclein".^[132] In 1919, Phoebus Levene identified the base, sugar and phosphate nucleotide unit.^[133] Levene suggested that DNA consisted of a string of nucleotide units linked together through the phosphate groups. However, Levene thought the chain was short and the bases repeated in a fixed order. In 1937 William Astbury produced the first X-ray diffraction patterns that showed that DNA had a regular structure.^[134]

In 1928, Frederick Griffith discovered that traits of the "smooth" form of the *Pneumococcus* could be transferred to the "rough" form of the same bacteria by mixing killed "smooth" bacteria with the live "rough" form.^[135] This system provided the first clear suggestion that DNA carried genetic information—the Avery-MacLeod-McCarty experiment—when Oswald Avery, along with coworkers Colin MacLeod and Maclyn McCarty, identified DNA as the transforming principle in 1943.^[136] DNA's role in heredity was confirmed in 1952, when Alfred Hershey and Martha Chase in the Hershey-Chase experiment showed that DNA is the genetic material of the T2 phage.^[137]



Francis Crick



Rosalind Franklin



Raymond Gosling

In 1953 James D. Watson and Francis Crick suggested what is now accepted as the first correct double-helix model of DNA structure in the journal *Nature*.^[7] Their double-helix, molecular model of DNA was then based on a single X-ray diffraction image (labeled as "Photo 51")^[138] taken by Rosalind Franklin and Raymond Gosling in May 1952, as well as the information that the DNA bases were paired—also obtained through private communications from Erwin Chargaff in the previous years. Chargaff's rules played a very important role in establishing double-helix configurations for B-DNA as well as A-DNA.

Experimental evidence supporting the Watson and Crick model were published in a series of five articles in the same issue of *Nature*.^[139] Of these, Franklin and Gosling's paper was the first publication of their own X-ray diffraction data and original analysis method that partially supported the Watson and Crick model^{[31] [140]}; this issue also contained an article on DNA structure by Maurice Wilkins and two of his colleagues, whose analysis and *in vivo* B-DNA X-ray patterns also supported the presence *in vivo* of the double-helical DNA configurations as proposed by Crick and Watson for their double-helix molecular model of DNA in the previous two pages of *Nature*.^[32] In 1962, after Franklin's death, Watson, Crick, and Wilkins jointly received the Nobel Prize in Physiology or Medicine.^[141] Unfortunately, Nobel rules of the time allowed only living recipients, but a vigorous debate continues on who should receive credit for the discovery.^[142]

In an influential presentation in 1957, Crick laid out the "Central Dogma" of molecular biology, which foretold the relationship between DNA, RNA, and proteins, and articulated the "adaptor hypothesis".^[143] Final confirmation of the replication mechanism that was implied by the double-helical structure followed in 1958 through the Meselson-Stahl experiment.^[144] Further work by Crick and coworkers showed that the genetic code was based on non-overlapping triplets of bases, called codons, allowing Har Gobind Khorana, Robert W. Holley and Marshall Warren Nirenberg to decipher the genetic code.^[145] These findings represent the birth of molecular biology.

See also

- Crystallography
- DNA microarray
- DNA sequencing
- Genetic disorder
- Junk DNA
- Molecular models of DNA
- Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid
- Nucleic acid analogues
- Nucleic acid methods
- Nucleic acid modeling
- Nucleic Acid Notations
- Paracrystal model and theory
- X-ray crystallography
- X-ray scattering
- Phosphoramidite
- Plasmid
- Polymerase chain reaction
- *Proteopedia DNA* ^[146]
- Southern blot
- Triple-stranded DNA

Further reading

- Calladine, Chris R.; Drew, Horace R.; Luisi, Ben F. and Travers, Andrew A. (2003). *Understanding DNA: the molecule & how it works*. Amsterdam: Elsevier Academic Press. ISBN 0-12-155089-3.
- Dennis, Carina; Julie Clayton (2003). *50 years of DNA*. Basingstoke: Palgrave Macmillan. ISBN 1-4039-1479-6.
- Judson, Horace Freeland (1996). *The eighth day of creation: makers of the revolution in biology*. Plainview, N.Y: CSHL Press. ISBN 0-87969-478-5.
- Olby, Robert C. (1994). *The path to the double helix: the discovery of DNA*. New York: Dover Publications. ISBN 0-486-68117-3., first published in October 1974 by MacMillan, with foreword by Francis Crick;the definitive DNA textbook, revised in 1994 with a 9 page postscript.
- Olby, Robert C. (2009). *Francis Crick: A Biography*. Plainview, N.Y: Cold Spring Harbor Laboratory Press. ISBN 0-87969-798-9.
- Ridley, Matt (2006). *Francis Crick: discoverer of the genetic code*. [Ashland, OH: Eminent Lives, Atlas Books. ISBN 0-06-082333-X.
- Berry, Andrew; Watson, James D. (2003). *DNA: the secret of life*. New York: Alfred A. Knopf. ISBN 0-375-41546-7.
- Stent, Gunther Siegmund; Watson, James D. (1980). *The double helix: a personal account of the discovery of the structure of DNA*. New York: Norton. ISBN 0-393-95075-1.

- Wilkins, Maurice (2003). *The third man of the double helix the autobiography of Maurice Wilkins*. Cambridge, Eng: University Press. ISBN 0-19-860665-6.

External links

- DNA ^[147] at the Open Directory Project
- DNA binding site prediction on protein ^[148]
- DNA coiling to form chromosomes ^[149]
- DNA from the Beginning ^[150] Another DNA Learning Center site on DNA, genes, and heredity from Mendel to the human genome project.
- DNA Lab, demonstrates how to extract DNA from wheat using readily available equipment and supplies. ^[151]
- DNA the Double Helix Game ^[152] From the official Nobel Prize web site
- DNA under electron microscope ^[153]
- Dolan DNA Learning Center ^[154]
- Double Helix: 50 years of DNA ^[155], *Nature*
- Double Helix 1953–2003 ^[156] National Centre for Biotechnology Education
- Francis Crick and James Watson talking on the BBC in 1962, 1972, and 1974 ^[157]
- Genetic Education Modules for Teachers ^[158]—*DNA from the Beginning* Study Guide
- Guide to DNA cloning ^[159]
- Olby R (January 2003). "Quiet debut for the double helix" ^[160]. *Nature* **421** (6921): 402–5. doi:10.1038/nature01397. PMID 12540907.
- PDB Molecule of the Month *pdb23_1* ^[161]
- Rosalind Franklin's contributions to the study of DNA ^[162]
- The Register of Francis Crick Personal Papers 1938 - 2007 ^[163] at Mandeville Special Collections Library, Geisel Library, University of California, San Diego
- U.S. National DNA Day ^[164]—watch videos and participate in real-time chat with top scientists
- "Clue to chemistry of heredity found" ^[165]. *The New York Times*. Saturday, June 13, 1953. The first American newspaper coverage of the discovery of the DNA structure.
- An Introduction to DNA and Chromosomes ^[166] from HOPES: Huntington's Disease Outreach Project for Education at Stanford

References

- [1] Russell, Peter (2001). *iGenetics*. New York: Benjamin Cummings. ISBN 0-805-34553-1.
- [2] Saenger, Wolfram (1984). *Principles of Nucleic Acid Structure*. New York: Springer-Verlag. ISBN 0387907629.
- [3] Alberts, Bruce; Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts and Peter Walters (2002). *Molecular Biology of the Cell; Fourth Edition* (<http://www.ncbi.nlm.nih.gov/books/bv.fcgi?call=bv.View..ShowTOC&rid=mboc4.TOC&depth=2>). New York and London: Garland Science. ISBN 0-8153-3218-1. OCLC 145080076 48122761 57023651 69932405. .
- [4] Butler, John M. (2001). *Forensic DNA Typing*. Elsevier. ISBN 978-0-12-147951-0. OCLC 223032110 45406517. pp. 14–15.
- [5] Mandelkern M, Elias J, Eden D, Crothers D (1981). "The dimensions of DNA in solution". *J Mol Biol* **152** (1): 153–61. doi:10.1016/0022-2836(81)90099-1. PMID 7338906.
- [6] Gregory S; Barlow, KF; McLay, KE; Kaul, R; Swarbreck, D; Dunham, A; Scott, CE; Howe, KL *et al.* (2006). "The DNA sequence and biological annotation of human chromosome 1". *Nature* **441** (7091): 315–21. doi:10.1038/nature04727. PMID 16710414.
- [7] Watson J.D. and Crick F.H.C. (1953). "A Structure for Deoxyribose Nucleic Acid" (<http://www.nature.com/nature/dna50/watsoncrick.pdf>) (PDF). *Nature* **171** (4356): 737–738. doi:10.1038/171737a0. PMID 13054692. . Retrieved 4 May 2009.
- [8] Berg J., Tymoczko J. and Stryer L. (2002) *Biochemistry*. W. H. Freeman and Company ISBN 0-7167-4955-6
- [9] Abbreviations and Symbols for Nucleic Acids, Polynucleotides and their Constituents (<http://www.chem.qmul.ac.uk/iupac/misc/naabb.html>) IUPAC-IUB Commission on Biochemical Nomenclature (CBN), Accessed 03 January 2006
- [10] Ghosh A, Bansal M (2003). "A glossary of DNA structures from A to Z". *Acta Crystallogr D Biol Crystallogr* **59** (Pt 4): 620–6. doi:10.1107/S0907444903003251. PMID 12657780.
- [11] Created from PDB 1D65 (<http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1D65>)

- [12] Verma S, Eckstein F (1998). "Modified oligonucleotides: synthesis and strategy for users". *Annu. Rev. Biochem.* **67**: 99–134. doi:10.1146/annurev.biochem.67.1.99. PMID 9759484.
- [13] Wing R, Drew H, Takano T, Broka C, Tanaka S, Itakura K, Dickerson R (1980). "Crystal structure analysis of a complete turn of B-DNA". *Nature* **287** (5784): 755–8. doi:10.1038/287755a0. PMID 7432492.
- [14] Pabo C, Sauer R (1984). "Protein-DNA recognition". *Annu Rev Biochem* **53**: 293–321. doi:10.1146/annurev.bi.53.070184.001453. PMID 6236744.
- [15] Clausen-Schaumann H, Rief M, Tolksdorf C, Gaub H (2000). "Mechanical stability of single DNA molecules" (<http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=1300792>). *Biophys J* **78** (4): 1997–2007. doi:10.1016/S0006-3495(00)76747-6. PMID 10733978. PMC 1300792.
- [16] Yakovchuk P, Protozanova E, Frank-Kamenetskii MD (2006). "Base-stacking and base-pairing contributions into thermal stability of the DNA double helix" (<http://nar.oxfordjournals.org/cgi/pmidlookup?view=long&pmid=16449200>). *Nucleic Acids Res.* **34** (2): 564–74. doi:10.1093/nar/gkj454. PMID 16449200. PMC 1360284. .
- [17] Chalikian T, Völker J, Plum G, Breslauer K (1999). "A more unified picture for the thermodynamics of nucleic acid duplex melting: a characterization by calorimetric and volumetric techniques" (<http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=22151>). *Proc Natl Acad Sci USA* **96** (14): 7853–8. doi:10.1073/pnas.96.14.7853. PMID 10393911. PMC 22151.
- [18] deHaseth P, Helmann J (1995). "Open complex formation by Escherichia coli RNA polymerase: the mechanism of polymerase-induced strand separation of double helical DNA". *Mol Microbiol* **16** (5): 817–24. doi:10.1111/j.1365-2958.1995.tb02309.x. PMID 7476180.
- [19] Isaksson J, Acharya S, Barman J, Cheruku P, Chattopadhyaya J (2004). "Single-stranded adenine-rich DNA and RNA retain structural characteristics of their respective double-stranded conformations and show directional differences in stacking pattern". *Biochemistry* **43** (51): 15996–6010. doi:10.1021/bi048221v. PMID 15609994.
- [20] Designation of the two strands of DNA (<http://www.chem.qmul.ac.uk/iubmb/newsletter/misc/DNA.html>) JCBN/NC-IUB Newsletter 1989, Accessed 07 May 2008
- [21] Hüttenhofer A, Schattner P, Polacek N (2005). "Non-coding RNAs: hope or hype?". *Trends Genet* **21** (5): 289–97. doi:10.1016/j.tig.2005.03.007. PMID 15851066.
- [22] Munroe S (2004). "Diversity of antisense regulation in eukaryotes: multiple mechanisms, emerging patterns". *J Cell Biochem* **93** (4): 664–71. doi:10.1002/jcb.20252. PMID 15389973.
- [23] Makalowska I, Lin C, Makalowski W (2005). "Overlapping genes in vertebrate genomes". *Comput Biol Chem* **29** (1): 1–12. doi:10.1016/j.compbiolchem.2004.12.006. PMID 15680581.
- [24] Johnson Z, Chisholm S (2004). "Properties of overlapping genes are conserved across microbial genomes" (<http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=525685>). *Genome Res* **14** (11): 2268–72. doi:10.1101/gr.2433104. PMID 15520290. PMC 525685.
- [25] Lamb R, Horvath C (1991). "Diversity of coding strategies in influenza viruses". *Trends Genet* **7** (8): 261–6. PMID 1771674.
- [26] Benham C, Mielke S (2005). "DNA mechanics". *Annu Rev Biomed Eng* **7**: 21–53. doi:10.1146/annurev.bioeng.6.062403.132016. PMID 16004565.
- [27] Champoux J (2001). "DNA topoisomerases: structure, function, and mechanism". *Annu Rev Biochem* **70**: 369–413. doi:10.1146/annurev.biochem.70.1.369. PMID 11395412.
- [28] Wang J (2002). "Cellular roles of DNA topoisomerases: a molecular perspective". *Nat Rev Mol Cell Biol* **3** (6): 430–40. doi:10.1038/nrm831. PMID 12042765.
- [29] Basu H, Feuerstein B, Zarlind D, Shafer R, Marton L (1988). "Recognition of Z-RNA and Z-DNA determinants by polyamines in solution: experimental and theoretical studies". *J Biomol Struct Dyn* **6** (2): 299–309. PMID 2482766.
- [30] Franklin RE, Gosling RG (6 March 1953). "The Structure of Sodium Thymonucleate Fibres I. The Influence of Water Content" (http://hekto.med.unc.edu:8080/CARTER/carter_WWW/Bioch_134/PDF_files/Franklin_Gosling.pdf). *Acta Cryst* **6** (8-9): 673–7. doi:10.1107/S0365110X53001939. .
Franklin RE, Gosling RG (September 1953). "The structure of sodium thymonucleate fibres. II. The cylindrically symmetrical Patterson function". *Acta Cryst* **6** (8-9): 678–85. doi:10.1107/S0365110X53001940.
- [31] Franklin, Rosalind and Gosling, Raymond (1953). "Molecular Configuration in Sodium Thymonucleate. Franklin R. and Gosling R.G" (<http://www.nature.com/nature/dna50/franklingosling.pdf>) (PDF). *Nature* **171** (4356): 740–1. doi:10.1038/171740a0. PMID 13054694. .
- [32] Wilkins M.H.F., A.R. Stokes A.R. & Wilson, H.R. (1953). "Molecular Structure of Deoxypentose Nucleic Acids" (<http://www.nature.com/nature/dna50/wilkins.pdf>) (PDF). *Nature* **171** (4356): 738–740. doi:10.1038/171738a0. PMID 13054693. .
- [33] Leslie AG, Arnott S, Chandrasekaran R, Ratliff RL (1980). "Polymorphism of DNA double helices". *J. Mol. Biol.* **143** (1): 49–72. doi:10.1016/0022-2836(80)90124-2. PMID 7441761.
- [34] Baianu, I.C. (1980). "Structural Order and Partial Disorder in Biological systems". *Bull. Math. Biol.* **42** (4): 137–141. <http://cogprints.org/3822/>
- [35] Hosemann R., Bagchi R.N., *Direct analysis of diffraction by matter*, North-Holland Publs., Amsterdam – New York, 1962.
- [36] Baianu, I.C. (1978). "X-ray scattering by partially disordered membrane systems.". *Acta Cryst.*, **A34** (5): 751–753. doi:10.1107/S0567739478001540.
- [37] Wahl M, Sundaralingam M (1997). "Crystal structures of A-DNA duplexes". *Biopolymers* **44** (1): 45–63. doi:10.1002/(SICI)1097-0282(1997)44:1 (inactive 2009-03-14). PMID 9097733.

- [38] Lu XJ, Shakked Z, Olson WK (2000). "A-form conformational motifs in ligand-bound DNA structures". *J. Mol. Biol.* **300** (4): 819–40. doi:10.1006/jmbi.2000.3690. PMID 10891271.
- [39] Rothenburg S, Koch-Nolte F, Haag F (2001). "DNA methylation and Z-DNA formation as mediators of quantitative differences in the expression of alleles". *Immunol Rev* **184**: 286–98. doi:10.1034/j.1600-065x.2001.1840125.x. PMID 12086319.
- [40] Oh D, Kim Y, Rich A (2002). "Z-DNA-binding proteins can act as potent effectors of gene expression in vivo" (<http://www.pnas.org/cgi/pmidlookup?view=long&pmid=12486233>). *Proc. Natl. Acad. Sci. U.S.A.* **99** (26): 16666–71. doi:10.1073/pnas.262672699. PMID 12486233. PMC 139201. .
- [41] Created from NDB UD0017 (<http://ndbserver.rutgers.edu/atlas/xray/structures/U/ud0017/ud0017.html>)
- [42] Greider C, Blackburn E (1985). "Identification of a specific telomere terminal transferase activity in Tetrahymena extracts". *Cell* **43** (2 Pt 1): 405–13. doi:10.1016/0092-8674(85)90170-9. PMID 3907856.
- [43] Nugent C, Lundblad V (1998). "The telomerase reverse transcriptase: components and regulation" (<http://www.genesdev.org/cgi/content/full/12/8/1073>). *Genes Dev* **12** (8): 1073–85. doi:10.1101/gad.12.8.1073. PMID 9553037. .
- [44] Wright W, Tesmer V, Huffman K, Levene S, Shay J (1997). "Normal human chromosomes have long G-rich telomeric overhangs at one end" (<http://www.genesdev.org/cgi/content/full/11/21/2801>). *Genes Dev* **11** (21): 2801–9. doi:10.1101/gad.11.21.2801. PMID 9353250. PMC 316649. .
- [45] Burge S, Parkinson G, Hazel P, Todd A, Neidle S (2006). "Quadruplex DNA: sequence, topology and structure" (<http://nar.oxfordjournals.org/cgi/pmidlookup?view=long&pmid=17012276>). *Nucleic Acids Res* **34** (19): 5402–15. doi:10.1093/nar/gkl655. PMID 17012276. PMC 1636468. .
- [46] Parkinson G, Lee M, Neidle S (2002). "Crystal structure of parallel quadruplexes from human telomeric DNA". *Nature* **417** (6891): 876–80. doi:10.1038/nature755. PMID 12050675.
- [47] Griffith J, Comeau L, Rosenfield S, Stansel R, Bianchi A, Moss H, de Lange T (1999). "Mammalian telomeres end in a large duplex loop". *Cell* **97** (4): 503–14. doi:10.1016/S0092-8674(00)80760-6. PMID 10338214.
- [48] Seeman NC (November 2005). "DNA enables nanoscale control of the structure of matter". *Q. Rev. Biophys.* **38** (4): 363–71. doi:10.1017/S0033583505004087. PMID 16515737.
- [49] Klose R, Bird A (2006). "Genomic DNA methylation: the mark and its mediators". *Trends Biochem Sci* **31** (2): 89–97. doi:10.1016/j.tibs.2005.12.008. PMID 16403636.
- [50] Bird A (2002). "DNA methylation patterns and epigenetic memory". *Genes Dev* **16** (1): 6–21. doi:10.1101/gad.947102. PMID 11782440.
- [51] Walsh C, Xu G (2006). "Cytosine methylation and DNA repair". *Curr Top Microbiol Immunol* **301**: 283–315. doi:10.1007/3-540-31390-7_11. PMID 16570853.
- [52] Kriaucionis S, Heintz N (May 2009). "The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain". *Science* **324** (5929): 929–30. doi:10.1126/science.1169786. PMID 19372393.
- [53] Ratel D, Ravanat J, Berger F, Wion D (2006). "N6-methyladenine: the other methylated base of DNA" (<http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=2754416>). *Bioessays* **28** (3): 309–15. doi:10.1002/bies.20342. PMID 16479578. PMC 2754416.
- [54] Gommers-Ampt J, Van Leeuwen F, de Beer A, Vliegthart J, Dizdaroğlu M, Kowalak J, Crain P, Borst P (1993). "beta-D-glucosyl-hydroxymethyluracil: a novel modified base present in the DNA of the parasitic protozoan *T. brucei*". *Cell* **75** (6): 1129–36. doi:10.1016/0092-8674(93)90322-H. PMID 8261512.
- [55] Created from PDB 1JDG (<http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1JDG>)
- [56] Douki T, Reynaud-Angelin A, Cadet J, Sage E (2003). "Bipyrimidine photoproducts rather than oxidative lesions are the main type of DNA damage involved in the genotoxic effect of solar UVA radiation". *Biochemistry* **42** (30): 9221–6. doi:10.1021/bi034593c. PMID 12885257. .
- [57] Cadet J, Delatour T, Douki T, Gasparutto D, Pouget J, Ravanat J, Sauvaigo S (1999). "Hydroxyl radicals and DNA base damage". *Mutat Res* **424** (1–2): 9–21. PMID 10064846.
- [58] Beckman KB, Ames BN (August 1997). "Oxidative decay of DNA" (<http://www.jbc.org/cgi/pmidlookup?view=long&pmid=9289489>). *J. Biol. Chem.* **272** (32): 19633–6. doi:10.1074/jbc.272.32.19633. PMID 9289489. .
- [59] Valerie K, Povirk L (2003). "Regulation and mechanisms of mammalian double-strand break repair". *Oncogene* **22** (37): 5792–812. doi:10.1038/sj.onc.1206679. PMID 12947387.
- [60] Ferguson L, Denny W (1991). "The genetic toxicology of acridines". *Mutat Res* **258** (2): 123–60. PMID 1881402.
- [61] Jeffrey A (1985). "DNA modification by chemical carcinogens". *Pharmacol Ther* **28** (2): 237–72. doi:10.1016/0163-7258(85)90013-0. PMID 3936066.
- [62] Stephens T, Bunde C, Fillmore B (2000). "Mechanism of action in thalidomide teratogenesis". *Biochem Pharmacol* **59** (12): 1489–99. doi:10.1016/S0006-2952(99)00388-3. PMID 10799645.
- [63] Braña M, Cacho M, Gradillas A, de Pascual-Teresa B, Ramos A (2001). "Intercalators as anticancer drugs". *Curr Pharm Des* **7** (17): 1745–80. doi:10.2174/1381612013397113. PMID 11562309.
- [64] Venter J, Adams, MD; Myers, EW; Li, PW; Mural, RJ; Sutton, GG; Smith, HO; Yandell, M *et al.* (2001). "The sequence of the human genome". *Science* **291** (5507): 1304–51. doi:10.1126/science.1058040. PMID 11181995.
- [65] Thanbichler M, Wang S, Shapiro L (2005). "The bacterial nucleoid: a highly organized and dynamic structure". *J Cell Biochem* **96** (3): 506–21. doi:10.1002/jcb.20519. PMID 15988757.
- [66] Wolfsberg T, McEntyre J, Schuler G (2001). "Guide to the draft human genome". *Nature* **409** (6822): 824–6. doi:10.1038/35057000. PMID 11236998.

- [67] Gregory T (2005). "The C-value enigma in plants and animals: a review of parallels and an appeal for partnership" (<http://aob.oxfordjournals.org/cgi/content/full/95/1/133>). *Ann Bot (Lond)* **95** (1): 133–46. doi:10.1093/aob/mci009. PMID 15596463. .
- [68] The ENCODE Project Consortium (2007). "Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project" (<http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=2212820>). *Nature* **447** (7146): 799–816. doi:10.1038/nature05874. PMID 17571346. PMC 2212820.
- [69] Created from PDB 1MSW (<http://www.rcsb.org/pdb/explore/explore.do?structureId=1MSW>)
- [70] Pidoux A, Allshire R (2005). "The role of heterochromatin in centromere function" (<http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=1569473>). *Philos Trans R Soc Lond B Biol Sci* **360** (1455): 569–79. doi:10.1098/rstb.2004.1611. PMID 15905142. PMC 1569473.
- [71] Harrison P, Hegyi H, Balasubramanian S, Luscombe N, Bertone P, Echols N, Johnson T, Gerstein M (2002). "Molecular fossils in the human genome: identification and analysis of the pseudogenes in chromosomes 21 and 22" (<http://www.genome.org/cgi/content/full/12/2/272>). *Genome Res* **12** (2): 272–80. doi:10.1101/gr.207102. PMID 11827946. PMC 155275. .
- [72] Harrison P, Gerstein M (2002). "Studying genomes through the aeons: protein families, pseudogenes and proteome evolution". *J Mol Biol* **318** (5): 1155–74. doi:10.1016/S0022-2836(02)00109-2. PMID 12083509.
- [73] Albà M (2001). "Replicative DNA polymerases" (<http://genomebiology.com/1465-6906/2/REVIEWS3002>). *Genome Biol* **2** (1): REVIEWS3002. doi:10.1186/gb-2001-2-1-reviews3002. PMID 11178285. PMC 150442. .
- [74] Sandman K, Pereira S, Reeve J (1998). "Diversity of prokaryotic chromosomal proteins and the origin of the nucleosome". *Cell Mol Life Sci* **54** (12): 1350–64. doi:10.1007/s000180050259. PMID 9893710.
- [75] Dame RT (2005). "The role of nucleoid-associated proteins in the organization and compaction of bacterial chromatin". *Mol. Microbiol.* **56** (4): 858–70. doi:10.1111/j.1365-2958.2005.04598.x. PMID 15853876.
- [76] Luger K, Mäder A, Richmond R, Sargent D, Richmond T (1997). "Crystal structure of the nucleosome core particle at 2.8 Å resolution". *Nature* **389** (6648): 251–60. doi:10.1038/38444. PMID 9305837.
- [77] Jenuwein T, Allis C (2001). "Translating the histone code". *Science* **293** (5532): 1074–80. doi:10.1126/science.1063127. PMID 11498575.
- [78] Ito T (2003). "Nucleosome assembly and remodelling". *Curr Top Microbiol Immunol* **274**: 1–22. PMID 12596902.
- [79] Thomas J (2001). "HMG1 and 2: architectural DNA-binding proteins". *Biochem Soc Trans* **29** (Pt 4): 395–401. doi:10.1042/BST0290395. PMID 11497996.
- [80] Grosschedl R, Giese K, Pagel J (1994). "HMG domain proteins: architectural elements in the assembly of nucleoprotein structures". *Trends Genet* **10** (3): 94–100. doi:10.1016/0168-9525(94)90232-1. PMID 8178371.
- [81] Iftode C, Daniely Y, Borowiec J (1999). "Replication protein A (RPA): the eukaryotic SSB". *Crit Rev Biochem Mol Biol* **34** (3): 141–80. doi:10.1080/10409239991209255. PMID 10473346.
- [82] Created from PDB 1LMB (<http://www.rcsb.org/pdb/explore/explore.do?structureId=1LMB>)
- [83] Myers L, Kornberg R (2000). "Mediator of transcriptional regulation". *Annu Rev Biochem* **69**: 729–49. doi:10.1146/annurev.biochem.69.1.729. PMID 10966474.
- [84] Spiegelman B, Heinrich R (2004). "Biological control through regulated transcriptional coactivators". *Cell* **119** (2): 157–67. doi:10.1016/j.cell.2004.09.037. PMID 15479634.
- [85] Li Z, Van Calcar S, Qu C, Cavenne W, Zhang M, Ren B (2003). "A global transcriptional regulatory role for c-Myc in Burkitt's lymphoma cells" (<http://www.pnas.org/cgi/pmidlookup?view=long&pmid=12808131>). *Proc Natl Acad Sci USA* **100** (14): 8164–9. doi:10.1073/pnas.1332764100. PMID 12808131. PMC 166200. .
- [86] Pabo C, Sauer R (1984). "Protein-DNA recognition". *Annu Rev Biochem* **53**: 293–321. doi:10.1146/annurev.bi.53.070184.001453. PMID 6236744.
- [87] Created from PDB 1RVA (<http://www.rcsb.org/pdb/explore/explore.do?structureId=1RVA>)
- [88] Bickle T, Krüger D (1993). "Biology of DNA restriction" (<http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=372918>). *Microbiol Rev* **57** (2): 434–50. PMID 8336674. PMC 372918.
- [89] Doherty A, Suh S (2000). "Structural and mechanistic conservation in DNA ligases" (<http://nar.oxfordjournals.org/cgi/pmidlookup?view=long&pmid=11058099>). *Nucleic Acids Res* **28** (21): 4051–8. doi:10.1093/nar/28.21.4051. PMID 11058099. PMC 113121. .
- [90] Schoeffler A, Berger J (2005). "Recent advances in understanding structure-function relationships in the type II topoisomerase mechanism". *Biochem Soc Trans* **33** (Pt 6): 1465–70. doi:10.1042/BST20051465. PMID 16246147.
- [91] Tuteja N, Tuteja R (2004). "Unraveling DNA helicases. Motif, structure, mechanism and function". *Eur J Biochem* **271** (10): 1849–63. doi:10.1111/j.1432-1033.2004.04094.x. PMID 15128295.
- [92] Joyce C, Steitz T (1995). "Polymerase structures and function: variations on a theme?" (<http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=177480>). *J Bacteriol* **177** (22): 6321–9. PMID 7592405. PMC 177480.
- [93] Hubscher U, Maga G, Spadari S (2002). "Eukaryotic DNA polymerases". *Annu Rev Biochem* **71**: 133–63. doi:10.1146/annurev.biochem.71.090501.150041. PMID 12045093.
- [94] Johnson A, O'Donnell M (2005). "Cellular DNA replicases: components and dynamics at the replication fork". *Annu Rev Biochem* **74**: 283–315. doi:10.1146/annurev.biochem.73.011303.073859. PMID 15952889.
- [95] Tarrago-Litvak L, Andréola M, Nevinsky G, Sarih-Cottin L, Litvak S (1 May 1994). "The reverse transcriptase of HIV-1: from enzymology to therapeutic intervention" (<http://www.fasebj.org/cgi/reprint/8/8/497>). *FASEB J* **8** (8): 497–503. PMID 7514143. .

- [96] Martinez E (2002). "Multi-protein complexes in eukaryotic gene transcription". *Plant Mol Biol* **50** (6): 925–47. doi:10.1023/A:1021258713850. PMID 12516863.
- [97] Created from PDB 1M6G (<http://www.rcsb.org/pdb/explore/explore.do?structureId=1M6G>)
- [98] Cremer T, Cremer C (2001). "Chromosome territories, nuclear architecture and gene regulation in mammalian cells". *Nat Rev Genet* **2** (4): 292–301. doi:10.1038/35066075. PMID 11283701.
- [99] Pál C, Papp B, Lercher M (2006). "An integrated view of protein evolution". *Nat Rev Genet* **7** (5): 337–48. doi:10.1038/nrg1838. PMID 16619049.
- [100] O'Driscoll M, Jeggo P (2006). "The role of double-strand break repair - insights from human genetics". *Nat Rev Genet* **7** (1): 45–54. doi:10.1038/nrg1746. PMID 16369571.
- [101] Vispé S, Defais M (1997). "Mammalian Rad51 protein: a RecA homologue with pleiotropic functions". *Biochimie* **79** (9-10): 587–92. doi:10.1016/S0300-9084(97)82007-X. PMID 9466696.
- [102] Neale MJ, Keeney S (2006). "Clarifying the mechanics of DNA strand exchange in meiotic recombination". *Nature* **442** (7099): 153–8. doi:10.1038/nature04885. PMID 16838012.
- [103] Dickman M, Ingleston S, Sedelnikova S, Rafferty J, Lloyd R, Grasby J, Hornby D (2002). "The RuvABC resolvosome". *Eur J Biochem* **269** (22): 5492–501. doi:10.1046/j.1432-1033.2002.03250.x. PMID 12423347.
- [104] Orgel L (2004). "Prebiotic chemistry and the origin of the RNA world" (<http://www.crbmb.com/cgi/reprint/39/2/99.pdf>) (PDF). *Crit Rev Biochem Mol Biol* **39** (2): 99–123. doi:10.1080/10409230490460765. PMID 15217990. .
- [105] Davenport R (2001). "Ribozymes. Making copies in the RNA world". *Science* **292** (5520): 1278. doi:10.1126/science.292.5520.1278a. PMID 11360970.
- [106] Szathmáry E (1992). "What is the optimum size for the genetic alphabet?" (<http://www.pnas.org/cgi/reprint/89/7/2614.pdf>) (PDF). *Proc Natl Acad Sci USA* **89** (7): 2614–8. doi:10.1073/pnas.89.7.2614. PMID 1372984. PMC 48712. .
- [107] Lindahl T (1993). "Instability and decay of the primary structure of DNA". *Nature* **362** (6422): 709–15. doi:10.1038/362709a0. PMID 8469282.
- [108] Vreeland R, Rosenzweig W, Powers D (2000). "Isolation of a 250 million-year-old halotolerant bacterium from a primary salt crystal". *Nature* **407** (6806): 897–900. doi:10.1038/35038060. PMID 11057666.
- [109] Hebsgaard M, Phillips M, Willerslev E (2005). "Geologically ancient DNA: fact or artefact?". *Trends Microbiol* **13** (5): 212–20. doi:10.1016/j.tim.2005.03.010. PMID 15866038.
- [110] Nickle D, Learn G, Rain M, Mullins J, Mittler J (2002). "Curiously modern DNA for a "250 million-year-old" bacterium". *J Mol Evol* **54** (1): 134–7. doi:10.1007/s00239-001-0025-x. PMID 11734907.
- [111] Goff SP, Berg P (1976). "Construction of hybrid viruses containing SV40 and lambda phage DNA segments and their propagation in cultured monkey cells". *Cell* **9** (4 PT 2): 695–705. doi:10.1016/0092-8674(76)90133-1. PMID 189942.
- [112] Houdebine L (2007). "Transgenic animal models in biomedical research". *Methods Mol Biol* **360**: 163–202. doi:10.1385/1-59745-165-7:163. PMID 17172731.
- [113] Daniell H, Dhingra A (2002). "Multigene engineering: dawn of an exciting new era in biotechnology". *Curr Opin Biotechnol* **13** (2): 136–41. doi:10.1016/S0958-1669(02)00297-5. PMID 11950565.
- [114] Job D (2002). "Plant biotechnology in agriculture". *Biochimie* **84** (11): 1105–10. doi:10.1016/S0300-9084(02)00013-5. PMID 12595138.
- [115] Collins A, Morton N (1994). "Likelihood ratios for DNA identification" (<http://www.pnas.org/cgi/reprint/91/13/6007.pdf>) (PDF). *Proc Natl Acad Sci USA* **91** (13): 6007–11. doi:10.1073/pnas.91.13.6007. PMID 8016106. PMC 44126. .
- [116] Weir B, Triggs C, Starling L, Stowell L, Walsh K, Buckleton J (1997). "Interpreting DNA mixtures". *J Forensic Sci* **42** (2): 213–22. PMID 9068179.
- [117] Jeffreys A, Wilson V, Thein S (1985). "Individual-specific 'fingerprints' of human DNA". *Nature* **316** (6023): 76–9. doi:10.1038/316076a0. PMID 2989708.
- [118] Colin Pitchfork — first murder conviction on DNA evidence also clears the prime suspect (http://www.forensic.gov.uk/forensic_t/inside/news/list_casefiles.php?case=1) Forensic Science Service Accessed 23 December 2006
- [119] "DNA Identification in Mass Fatality Incidents" (<http://massfatality.dna.gov/Introduction/>). National Institute of Justice. September 2006. .
- [120] Baldi, Pierre; Brunak, Soren (2001). *Bioinformatics: The Machine Learning Approach*. MIT Press. ISBN 978-0-262-02506-5. OCLC 45951728..
- [121] Gusfield, Dan. *Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology*. Cambridge University Press, 15 January 1997. ISBN 978-0-521-58519-4.
- [122] Sjölander K (2004). "Phylogenomic inference of protein molecular function: advances and challenges" (<http://bioinformatics.oxfordjournals.org/cgi/reprint/20/2/170>). *Bioinformatics* **20** (2): 170–9. doi:10.1093/bioinformatics/bth021. PMID 14734307. .
- [123] Mount DM (2004). *Bioinformatics: Sequence and Genome Analysis* (2 ed.). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press. ISBN 0879697121. OCLC 55106399.
- [124] Rothmund PW (March 2006). "Folding DNA to create nanoscale shapes and patterns". *Nature* **440** (7082): 297–302. doi:10.1038/nature04586. PMID 16541064.
- [125] Andersen ES, Dong M, Nielsen MM (May 2009). "Self-assembly of a nanoscale DNA box with a controllable lid". *Nature* **459** (7243): 73–6. doi:10.1038/nature07971. PMID 19424153.

- [126] Ishitsuka Y, Ha T (May 2009). "DNA nanotechnology: a nanomachine goes live". *Nat Nanotechnol* **4** (5): 281–2. doi:10.1038/nnano.2009.101. PMID 19421208.
- [127] Aldaye FA, Palmer AL, Sleiman HF (September 2008). "Assembling materials with DNA as the guide". *Science* **321** (5897): 1795–9. doi:10.1126/science.1154533. PMID 18818351.
- [128] Wray G; Martindale, Mark Q. (2002). "Dating branches on the tree of life using DNA" (<http://genomebiology.com/1465-6906/3/REVIEWS0001>). *Genome Biol* **3** (1): REVIEWS0001. doi:10.1046/j.1525-142X.1999.99010.x. PMID 11806830. PMC 150454. .
- [129] *Lost Tribes of Israel*, NOVA, PBS airdate: 22 February 2000. Transcript available from PBS.org, (<http://www.pbs.org/wgbh/nova/transcripts/2706israel.html>) (last accessed on 4 March 2006)
- [130] Kleiman, Yaakov. "The Cohanim/DNA Connection: The fascinating story of how DNA studies confirm an ancient biblical tradition". (http://www.aish.com/societywork/sciencenature/the_cohanim_-_dna_connection.asp) *aish.com* (January 13, 2000). Accessed 4 March 2006.
- [131] Bhattacharya, Shaoni. "Killer convicted thanks to relative's DNA". (<http://www.newscientist.com/article.ns?id=dn4908>) *newscientist.com* (20 April 2004). Accessed 22 December 06
- [132] Dahm R (January 2008). "Discovering DNA: Friedrich Miescher and the early years of nucleic acid research". *Hum. Genet.* **122** (6): 565–81. doi:10.1007/s00439-007-0433-0. PMID 17901982.
- [133] Levene P, (1 December 1919). "The structure of yeast nucleic acid" (<http://www.jbc.org/cgi/reprint/40/2/415>). *J Biol Chem* **40** (2): 415–24. .
- [134] Astbury W, (1947). "Nucleic acid". *Symp. SOC. Exp. Bbl* **1** (66).
- [135] Lorenz MG, Wackernagel W (1 September 1994). "Bacterial gene transfer by natural genetic transformation in the environment" (<http://mmbr.asm.org/cgi/pmidlookup?view=long&pmid=7968924>). *Microbiol. Rev.* **58** (3): 563–602. PMID 7968924. PMC 372978. .
- [136] Avery O, MacLeod C, McCarty M (1944). "Studies on the chemical nature of the substance inducing transformation of pneumococcal types. Inductions of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III" (<http://www.jem.org/cgi/reprint/149/2/297>). *J Exp Med* **79** (2): 137–158. doi:10.1084/jem.79.2.137. PMID 19871359. PMC 2135445. .
- [137] Hershey A, Chase M (1952). "Independent functions of viral protein and nucleic acid in growth of bacteriophage" (<http://www.jgp.org/cgi/reprint/36/1/39.pdf>) (PDF). *J Gen Physiol* **36** (1): 39–56. doi:10.1085/jgp.36.1.39. PMID 12981234. PMC 2147348. .
- [138] The B-DNA X-ray pattern on the right of this linked image (<http://osulibrary.oregonstate.edu/specialcollections/coll/pauling/dna/pictures/sci9.001.5.html>) was obtained by Rosalind Franklin and Raymond Gosling in May 1952 at high hydration levels of DNA and it has been labeled as "Photo 51"
- [139] Nature Archives Double Helix of DNA: 50 Years (<http://www.nature.com/nature/dna50/archive.html>)
- [140] Original X-ray diffraction image (<http://osulibrary.oregonstate.edu/specialcollections/coll/pauling/dna/pictures/franklin-typeBphoto.html>)
- [141] The Nobel Prize in Physiology or Medicine 1962 (http://nobelprize.org/nobel_prizes/medicine/laureates/1962/) Nobelprize .org Accessed 22 December 06
- [142] Brenda Maddox (23 January 2003). "The double helix and the 'wronged heroine'" (http://www.biomath.nyu.edu/index/course/hw_articles/nature4.pdf) (PDF). *Nature* **421** (6921): 407–408. doi:10.1038/nature01399. PMID 12540909. .
- [143] Crick, F.H.C. On degenerate templates and the adaptor hypothesis (PDF). (<http://genome.wellcome.ac.uk/assets/wtx030893.pdf>) genome.wellcome.ac.uk (Lecture, 1955). Accessed 22 December 2006
- [144] Meselson M, Stahl F (1958). "The replication of DNA in *Escherichia coli*" (<http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=528642>). *Proc Natl Acad Sci USA* **44** (7): 671–82. doi:10.1073/pnas.44.7.671. PMID 16590258. PMC 528642.
- [145] The Nobel Prize in Physiology or Medicine 1968 (http://nobelprize.org/nobel_prizes/medicine/laureates/1968/) Nobelprize.org Accessed 22 December 06
- [146] <http://proteopedia.org/wiki/index.php/DNA>
- [147] http://www.dmoz.org/Science/Biology/Biochemistry_and_Molecular_Biology/Biomolecules/Nucleic_Acids/DNA/
- [148] <http://pipe.scs.fsu.edu/displar.html>
- [149] http://biostudio.com/c_%20education%20mac.htm
- [150] <http://www.dnaftb.org/dnaftb/>
- [151] <http://ca.youtube.com/watch?v=iyb7fwduuGM>
- [152] http://nobelprize.org/educational_games/medicine/dna_double_helix/
- [153] http://www.fidelitysystems.com/Unlinked_DNA.html
- [154] <http://www.dnalc.org/>
- [155] <http://www.nature.com/nature/dna50/archive.html>
- [156] <http://www.ncbe.reading.ac.uk/DNA50/>
- [157] <http://www.bbc.co.uk/bbcfour/audiointerviews/profilepages/crickwatson1.shtml>
- [158] <http://www.genome.gov/10506718>
- [159] <http://www.blackwellpublishing.com/trun/artwork/Animations/cloningexp/cloningexp.html>
- [160] <http://chem-faculty.ucsd.edu/joseph/CHEM13/DNA1.pdf>
- [161] http://www.rcsb.org/pdb/static.do?p=education_discussion/molecule_of_the_month/pdb23_1.html
- [162] <http://mason.gmu.edu/~emoody/rfranklin.html>
- [163] <http://orpheus.ucsd.edu/speccoll/testing/html/mss0660a.html#abstract>

[164] <http://www.genome.gov/10506367>

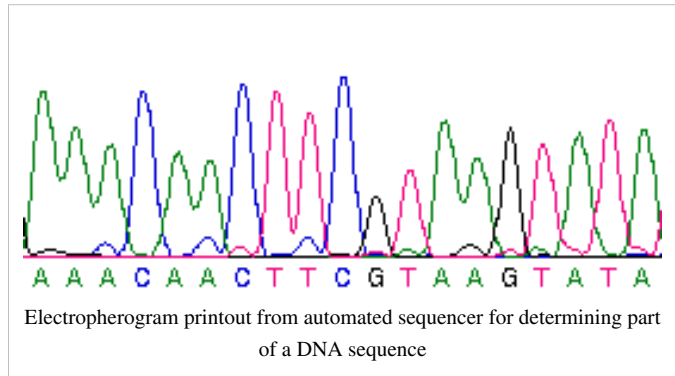
[165] <http://www.nytimes.com/packages/pdf/science/dna-article.pdf>

[166] <http://hopes.stanford.edu/basics/dna/b0.html>

DNA sequence

A **DNA sequence** or **genetic sequence** is a succession of letters representing the primary structure of a real or hypothetical DNA molecule or strand. The sequence has capacity to carry information. Sequences can be read from the biological raw material through DNA sequencing methods.

The possible letters are *A*, *C*, *G*, and *T*, representing the four nucleotide bases of a DNA strand — adenine, cytosine, guanine, thymine — covalently linked to a phosphodiester backbone. In the typical case, the sequences are printed abutting one another without gaps, as in the sequence AAAGTCTGAC, read left to right in the 5' to 3' direction. With regards to transcription, a sequence is on the coding strand if it has the same order as the transcribed RNA.



One sequence can be complementary to another sequence, meaning that they have the base on each position is the complementary (i.e. A to T, C to G) and in the reverse order. For example, the complementary sequence to TTAC is GTAA. If one strand of the double-stranded DNA is considered the sense strand, then the other strand, considered the antisense strand, will have the complementary sequence to the sense strand.

In some special cases, letters besides A, T, C, and G are present in a sequence. These letters represent ambiguity. Of all the molecules sampled, there is more than one kind of nucleotide at that position. The rules of the International Union of Pure and Applied Chemistry (IUPAC) are as follows:^[1]

- **A** = adenine
- **C** = cytosine
- **G** = guanine
- **T** = thymine
- **R** = G A (purine)
- **Y** = T C (pyrimidine)
- **K** = G T (keto)
- **M** = A C (amino)
- **S** = G C (strong bonds)
- **W** = A T (weak bonds)
- **B** = G T C (all but A)
- **D** = G A T (all but C)
- **H** = A C T (all but G)
- **V** = G C A (all but T)
- **N** = A G C T (any)

See also

- DNA
- DNA sequencing
- Single-nucleotide polymorphism (SNP)
- Sequence alignment
- Sequence analysis
- Sequence motif

External links

- A bibliography on features, patterns, correlations in DNA and protein texts ^[2]

References

- [1] Nomenclature for Incompletely Specified Bases in Nucleic Acid Sequences (<http://www.chem.qmul.ac.uk/iubmb/misc/naseq.html>), NC-IUB, 1984.
- [2] <http://www.nslj-genetics.org/dnacorr/>

DNA sequencing

The term **DNA sequencing** refers to sequencing methods for determining the order of the nucleotide bases—adenine, guanine, cytosine, and thymine—in a molecule of DNA.

Knowledge of DNA sequences has become indispensable for basic biological research, other research branches utilizing DNA sequencing, and in numerous applied fields such as diagnostic, biotechnology, forensic biology and biological systematics. The advent of DNA sequencing has significantly accelerated biological research and discovery. The rapid speed of sequencing attained with modern DNA sequencing technology has been instrumental in the sequencing of the human genome, in the Human Genome Project. Related projects, often by scientific collaboration across continents, have generated the complete DNA sequences of many animal, plant, and microbial genomes.

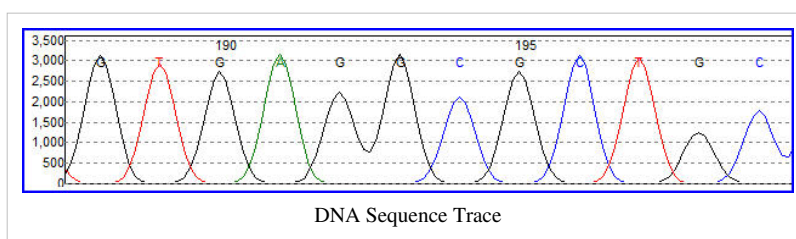
The first DNA sequences were obtained in the early 1970s by academic researchers using laborious methods based on two-dimensional chromatography. Following the development of dye-based sequencing methods with automated analysis,^[1] DNA sequencing has become easier and orders of magnitude faster.

History

RNA sequencing was one of the earliest forms of nucleotide sequencing. The major landmark of RNA sequencing is the sequence of the first complete gene and the complete

genome of Bacteriophage MS2, identified and published by Walter Fiers and his coworkers at the University of Ghent (Ghent, Belgium), between 1972^[2] and 1976.^[3]

Prior to the development of rapid DNA sequencing methods in the early 1970s by Frederick Sanger at the University of Cambridge, in England and Walter Gilbert and Allan Maxam at Harvard,^[4] ^[5] a number of laborious methods were used. For instance, in 1973, Gilbert and Maxam reported the sequence of 24 basepairs using a method known as wandering-spot analysis.^[6]



The chain-termination method developed by Sanger and coworkers in 1975 soon became the method of choice, owing to its relative ease and reliability.^{[7] [8]}

Maxam–Gilbert sequencing

In 1976–1977, Allan Maxam and Walter Gilbert developed a DNA sequencing method based on chemical modification of DNA and subsequent cleavage at specific bases.^[4] Although Maxam and Gilbert published their chemical sequencing method two years after the ground-breaking paper of Sanger and Coulson on plus-minus sequencing,^{[7] [9]} Maxam–Gilbert sequencing rapidly became more popular, since purified DNA could be used directly, while the initial Sanger method required that each read start be cloned for production of single-stranded DNA. However, with the improvement of the chain-termination method (see below), Maxam–Gilbert sequencing has fallen out of favour due to its technical complexity prohibiting its use in standard molecular biology kits, extensive use of hazardous chemicals, and difficulties with scale-up.

The method requires radioactive labelling at one end and purification of the DNA fragment to be sequenced. Chemical treatment generates breaks at a small proportion of one or two of the four nucleotide bases in each of four reactions (G, A+G, C, C+T). Thus a series of labelled fragments is generated, from the radiolabelled end to the first "cut" site in each molecule. The fragments in the four reactions are arranged side by side in gel electrophoresis for size separation. To visualize the fragments, the gel is exposed to X-ray film for autoradiography, yielding a series of dark bands each corresponding to a radiolabelled DNA fragment, from which the sequence may be inferred.

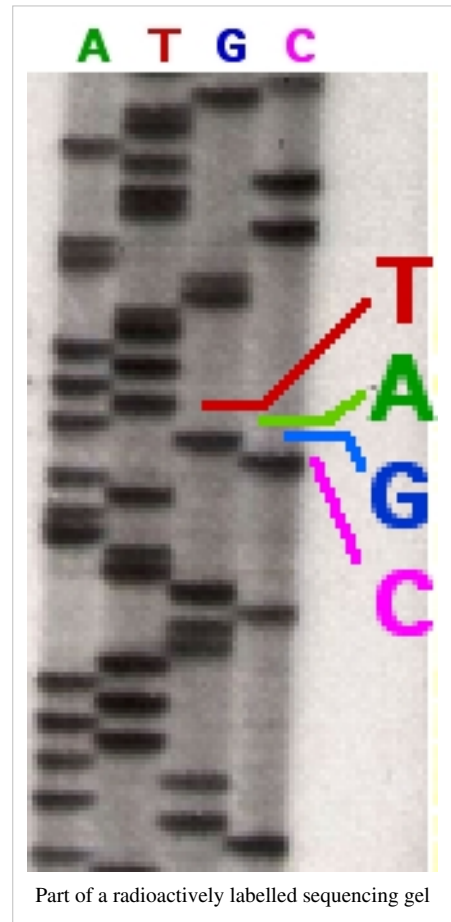
Also sometimes known as "chemical sequencing", this method originated in the study of DNA-protein interactions (footprinting), nucleic acid structure and epigenetic modifications to DNA, and within these it still has important applications.

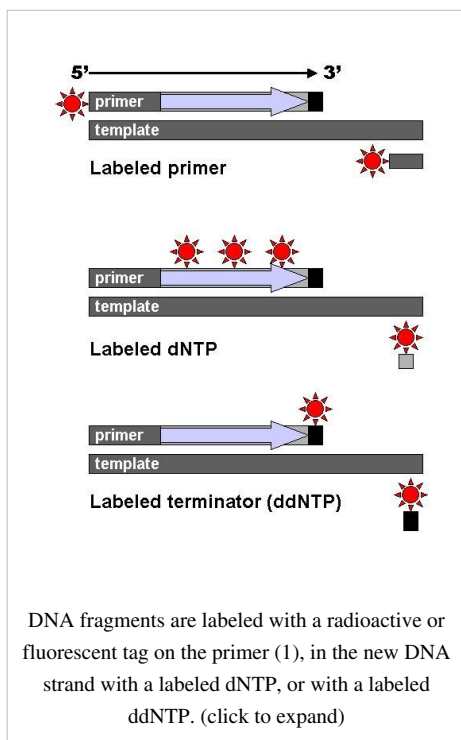
Chain-termination methods

Because the chain-terminator method (or Sanger method after its developer Frederick Sanger) is more efficient and uses fewer toxic chemicals and lower amounts of radioactivity than the method of Maxam and Gilbert, it rapidly became the method of choice. The key principle of the Sanger method was the use of dideoxynucleotide triphosphates (ddNTPs) as DNA chain terminators.

The classical chain-termination method requires a single-stranded DNA template, a DNA primer, a DNA polymerase, radioactively or fluorescently labeled nucleotides, and modified nucleotides that terminate DNA strand elongation. The DNA sample is divided into four separate sequencing reactions, containing all four of the standard deoxynucleotides (dATP, dGTP, dCTP and dTTP) and the DNA polymerase. To each reaction is added only one of the four dideoxynucleotides (ddATP, ddGTP, ddCTP, or ddTTP) which are the chain-terminating nucleotides, lacking a 3'-OH group required for the formation of a phosphodiester bond between two nucleotides, thus terminating DNA strand extension and resulting in DNA fragments of varying length.

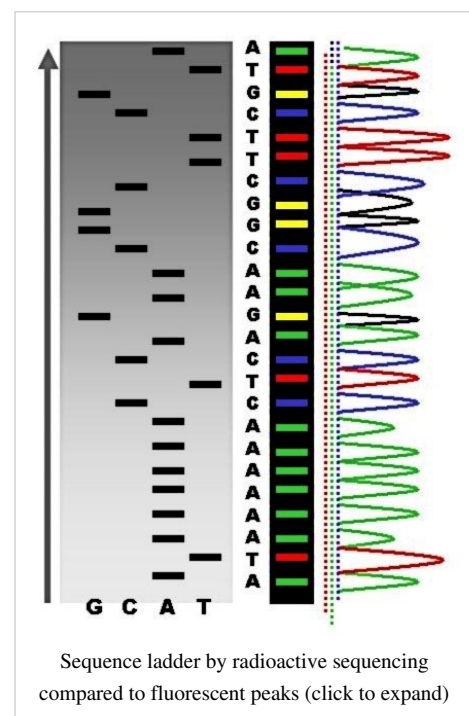
The newly synthesized and labeled DNA fragments are heat denatured, and separated by size (with a resolution of just one nucleotide) by gel electrophoresis on a denaturing polyacrylamide-urea gel with each of the four reactions run in one of four individual lanes (lanes A, T, G, C); the DNA bands are then visualized by autoradiography or UV light, and the DNA sequence can be directly read off the X-ray film or gel image. In the image on the right, X-ray film was exposed to the gel, and the dark bands correspond to DNA fragments of different lengths. A dark band in a lane indicates a DNA fragment that is the result of chain termination after incorporation of a dideoxynucleotide (ddATP, ddGTP, ddCTP, or ddTTP). The relative positions of the different bands among the four lanes are then used to read (from bottom to top) the DNA sequence.





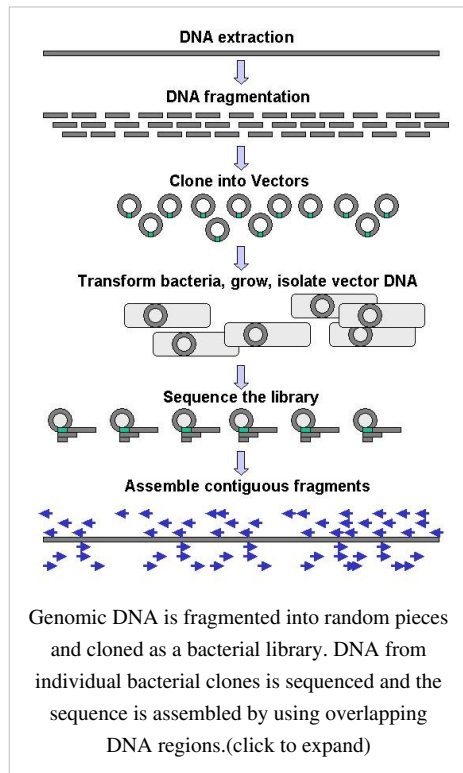
Technical variations of chain-termination sequencing include tagging with nucleotides containing radioactive phosphorus for radiolabelling, or using a primer labeled at the 5' end with a fluorescent dye. Dye-primer sequencing facilitates reading in an optical system for faster and more economical analysis and automation. The later development by Leroy Hood and coworkers^{[10] [11]} of fluorescently labeled ddNTPs and primers set the stage for automated, high-throughput DNA sequencing.

Chain-termination methods have greatly simplified DNA sequencing. For example, chain-termination-based kits are commercially available that contain the reagents needed for sequencing, pre-aliquoted and ready to use. Limitations include non-specific binding of the primer to the DNA, affecting accurate read-out of the DNA sequence, and DNA secondary structures affecting the fidelity of the sequence.



Large-scale sequencing strategies

Current methods can directly sequence only relatively short (300–1000 nucleotides long) DNA fragments in a single reaction.^[12] The main obstacle to sequencing DNA fragments above this size limit is insufficient power of separation for resolving large DNA fragments that differ in length by only one nucleotide.



Large-scale sequencing aims at sequencing very long DNA pieces, such as whole chromosomes. Common approaches consist of cutting (with restriction enzymes) or shearing (with mechanical forces) large DNA fragments into shorter DNA fragments. The fragmented DNA is cloned into a DNA vector, and amplified in *Escherichia coli*. Short DNA fragments purified from individual bacterial colonies are individually sequenced and assembled electronically into one long, contiguous sequence. This method does not require any pre-existing information about the sequence of the DNA and is referred to as *de novo* sequencing. Gaps in the assembled sequence may be filled by primer walking. The different strategies have different tradeoffs in speed and accuracy; *shotgun methods* are often used for sequencing large genomes, but its assembly is complex and difficult, particularly with sequence repeats often causing gaps in genome assembly.

New sequencing methods

High-throughput sequencing

The high demand for low-cost sequencing has driven the development of high-throughput sequencing technologies that parallelize the sequencing process, producing thousands or millions of sequences at once.^{[13] [14]} High-throughput sequencing technologies are intended to lower the cost of DNA sequencing beyond what is possible with standard dye-terminator methods.^[15]

In vitro clonal amplification

Most sequencing approaches use an *in vitro* cloning step to amplify individual DNA molecules, because their molecular detection methods are not sensitive enough for single molecule sequencing. Emulsion PCR isolates individual DNA molecules along with primer-coated beads in aqueous droplets within an oil phase. Polymerase chain reaction (PCR) then coats each bead with clonal copies of the DNA molecule followed by immobilization for later sequencing. Emulsion PCR is used in the methods by Margulis et al. (commercialized by 454 Life Sciences), Shendure and Porreca et al. (also known as "Polony sequencing") and SOLiD sequencing, (developed by Agencourt, now Applied Biosystems).^{[16] [17] [18]} Another method for *in vitro* clonal amplification is *bridge PCR*, where fragments are amplified upon primers attached to a solid surface, used in the Illumina Genome Analyzer. The single-molecule method developed by Stephen Quake's laboratory (later commercialized by Helicos) is an exception: it uses bright fluorophores and laser excitation to detect pyrosequencing events from individual DNA molecules fixed to a surface, eliminating the need for molecular amplification.^[19]

Parallelized sequencing

DNA molecules are physically bound to a surface, and sequenced in parallel. *Sequencing by synthesis*, like dye-termination electrophoretic sequencing, uses a DNA polymerase to determine the base sequence. Reversible terminator methods (used by Illumina and Helicos) use reversible versions of dye-terminators, adding one nucleotide at a time, detect fluorescence at each position in real time, by repeated removal of the blocking group to allow polymerization of another nucleotide. Pyrosequencing (used by 454) also uses DNA polymerization, adding one nucleotide species at a time and detecting and quantifying the number of nucleotides added to a given location through the light emitted by the release of attached pyrophosphates.^{[16] [20]}

Sequencing by ligation

Sequencing by ligation uses a DNA ligase to determine the target sequence.^{[17] [18] [21]} Used in the polony method and in the SOLiD technology, it uses a pool of all possible oligonucleotides of a fixed length, labeled according to the sequenced position. Oligonucleotides are annealed and ligated; the preferential ligation by DNA ligase for matching sequences results in a signal informative of the nucleotide at that position.

Microfluidic Sanger sequencing

In microfluidic Sanger sequencing the entire thermocycling amplification of DNA fragments as well as their separation by electrophoresis is done on a single glass wafer (approximately 10 cm in diameter) thus reducing the reagent usage as well as cost. In some instances researchers have shown that they can increase the throughput of conventional sequencing through the use of microchips. Research will still need to be done in order to make this use of technology effective.

Other sequencing technologies

Sequencing by hybridization is a non-enzymatic method that uses a DNA microarray. A single pool of DNA whose sequence is to be determined is fluorescently labeled and hybridized to an array containing known sequences. Strong hybridization signals from a given spot on the array identifies its sequence in the DNA being sequenced.^[22] Mass spectrometry may be used to determine mass differences between DNA fragments produced in chain-termination reactions.^[23]

DNA sequencing methods currently under development include labeling the DNA polymerase,^[24] reading the sequence as a DNA strand transits through nanopores,^{[25] [26]} and microscopy-based techniques, such as AFM or electron microscopy that are used to identify the positions of individual nucleotides within long DNA fragments (>5,000 bp) by nucleotide labeling with heavier elements (e.g., halogens) for visual detection and recording.^[27]

In October 2006, the X Prize Foundation established an initiative to promote the development of full genome sequencing technologies, called the Archon X Prize, intending to award \$10 million to "the first Team that can build a device and use it to sequence 100 human genomes within 10 days or less, with an accuracy of no more than one error in every 100,000 bases sequenced, with sequences accurately covering at least 98% of the genome, and at a recurring cost of no more than \$10,000 (US) per genome."^[28]

Major landmarks in DNA sequencing

- 1953 Discovery of the structure of the DNA double helix.
- 1972 Development of recombinant DNA technology, which permits isolation of defined fragments of DNA; prior to this, the only accessible samples for sequencing were from bacteriophage or virus DNA.
- 1975 The first complete DNA genome to be sequenced is that of bacteriophage ϕ X174
- 1977 Allan Maxam and Walter Gilbert publish "DNA sequencing by chemical degradation".^[4] Frederick Sanger, independently, publishes "DNA sequencing by enzymatic synthesis".
- 1980 Frederick Sanger and Walter Gilbert receive the Nobel Prize in Chemistry
- 1984 Medical Research Council scientists decipher the complete DNA sequence of the Epstein-Barr virus, 170 kb.
- 1986 Leroy E. Hood's laboratory at the California Institute of Technology and Smith announce the first semi-automated DNA sequencing machine.
- 1987 Applied Biosystems markets first automated sequencing machine, the model ABI 370.
- 1990 The U.S. National Institutes of Health (NIH) begins large-scale sequencing trials on *Mycoplasma capricolum*, *Escherichia coli*, *Caenorhabditis elegans*, and *Saccharomyces cerevisiae* (at 75 cents (US)/base).
- 1995 Craig Venter, Hamilton Smith, and colleagues at The Institute for Genomic Research (TIGR) publish the first complete genome of a free-living organism, the bacterium *Haemophilus influenzae*. The circular chromosome contains 1,830,137 bases and its publication in the journal *Science*^[29] marks the first use of whole-genome shotgun sequencing, eliminating the need for initial mapping efforts.
- 1995 Richard Mathies *et al.* publish fluorescence energy transfer dye-based sequencing.^[30]
- 1996 Pål Nyrén and his student Mostafa Ronaghi at the Royal Institute of Technology in Stockholm publish their method of pyrosequencing^[20]
- 1998 Phil Green and Brent Ewing of the University of Washington publish "phred" for sequencer data analysis.^[31]

See also

- Full genome sequencing
 - Genome project
 - Cancer genome sequencing
 - Single Molecule Real Time Sequencing
 - Applied Biosystems
 - 454 Life Sciences
 - Illumina (company)
 - Complete Genomics
 - Joint Genome Institute
 - DNA field-effect transistor
 - DNA sequencing theory
-

External links

- Disruptive Gene Sequencing technology - Single Molecule Real Time (SMRT) sequencing

References

- [1] Olsvik O, Wahlberg J, Petterson B, *et al.* (January 1993). "Use of automated sequencing of polymerase chain reaction-generated amplicons to identify three types of cholera toxin subunit B in *Vibrio cholerae* O1 strains" (<http://jcm.asm.org/cgi/pmidlookup?view=long&pmid=7678018>). *J. Clin. Microbiol.* **31** (1): 22–5. PMID 7678018. PMC 262614. .
- [2] Min Jou W, Haegeman G, Ysebaert M, Fiers W (May 1972). "Nucleotide sequence of the gene coding for the bacteriophage MS2 coat protein". *Nature* **237** (5350): 82–8. doi:10.1038/237082a0. PMID 4555447.
- [3] Fiers W, Contreras R, Duerinck F, *et al* (April 1976). "Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene". *Nature* **260** (5551): 500–7. doi:10.1038/260500a0. PMID 1264203.
- [4] Maxam AM, Gilbert W (February 1977). "A new method for sequencing DNA" (<http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=392330>). *Proc. Natl. Acad. Sci. U.S.A.* **74** (2): 560–4. doi:10.1073/pnas.74.2.560. PMID 265521. PMC 392330.
- [5] Gilbert, W. DNA sequencing and gene structure (http://nobelprize.org/nobel_prizes/chemistry/laureates/1980/gilbert-lecture.pdf). Nobel lecture, 8 December 1980.
- [6] Gilbert W, Maxam A (December 1973). "The nucleotide sequence of the lac operator" (<http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=427284>). *Proc. Natl. Acad. Sci. U.S.A.* **70** (12): 3581–4. doi:10.1073/pnas.70.12.3581. PMID 4587255. PMC 427284.
- [7] Sanger F, Coulson AR (May 1975). "A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase". *J. Mol. Biol.* **94** (3): 441–8. doi:10.1016/0022-2836(75)90213-2. PMID 1100841.
- [8] Sanger F, Nicklen S, Coulson AR (December 1977). "DNA sequencing with chain-terminating inhibitors" (<http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=431765>). *Proc. Natl. Acad. Sci. U.S.A.* **74** (12): 5463–7. doi:10.1073/pnas.74.12.5463. PMID 271968. PMC 431765.
- [9] Sanger F. Determination of nucleotide sequences in DNA (http://nobelprize.org/nobel_prizes/chemistry/laureates/1980/sanger-lecture.pdf). Nobel lecture, 8 December 1980.
- [10] Smith LM, Sanders JZ, Kaiser RJ, *et al* (1986). "Fluorescence detection in automated DNA sequence analysis". *Nature* **321** (6071): 674–9. doi:10.1038/321674a0. PMID 3713851. "We have developed a method for the partial automation of DNA sequence analysis. Fluorescence detection of the DNA fragments is accomplished by means of a fluorophore covalently attached to the oligonucleotide primer used in enzymatic DNA sequence analysis. A different coloured fluorophore is used for each of the reactions specific for the bases A, C, G and T. The reaction mixtures are combined and co-electrophoresed down a single polyacrylamide gel tube, the separated fluorescent bands of DNA are detected near the bottom of the tube, and the sequence information is acquired directly by computer."
- [11] Smith LM, Fung S, Hunkapiller MW, Hunkapiller TJ, Hood LE (April 1985). "The synthesis of oligonucleotides containing an aliphatic amino group at the 5' terminus: synthesis of fluorescent DNA primers for use in DNA sequence analysis" (<http://nar.oxfordjournals.org/cgi/pmidlookup?view=long&pmid=4000959>). *Nucleic Acids Res.* **13** (7): 2399–412. doi:10.1093/nar/13.7.2399. PMID 4000959. PMC 341163. .
- [12] 3730x1 DNA Analyzer (<http://www.appliedbiosystems.com/catalog/myab/StoreCatalog/products/CategoryDetails.jsp?hierarchyID=102&category1st=a50&category2nd=a51&category3rd=111907>)
- [13] Hall N (May 2007). "Advanced sequencing technologies and their wider impact in microbiology". *J. Exp. Biol.* **210** (Pt 9): 1518–25. doi:10.1242/jeb.001370. PMID 17449817.
- [14] Church GM (January 2006). "Genomes for all". *Sci. Am.* **294** (1): 46–54. doi:10.1038/scientificamerican0106-46. PMID 16468433.
- [15] Schuster, Stephan C. (2008). *Next-generation sequencing transforms today's biology*. **5**. Nature Methods. pp. 16–18. doi:10.1038/nmeth1156.
- [16] Margulies M, Egholm M, Altman WE, *et al* (September 2005). "Genome sequencing in microfabricated high-density picolitre reactors" (<http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=1464427>). *Nature* **437** (7057): 376–80. doi:10.1038/nature03959. PMID 16056220. PMC 1464427.
- [17] Shendure, J. (2005). "Accurate Multiplex Polony Sequencing of an Evolved Bacterial Genome". *Science* **309**: 1728. doi:10.1126/science.1117389.
- [18] Applied Biosystems' SOLiD technology (<http://solid.appliedbiosystems.com/>)
- [19] Braslavsky I, Hebert B, Kartalov E, Quake SR (April 2003). "Sequence information can be obtained from single DNA molecules" (<http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=153030>). *Proc. Natl. Acad. Sci. U.S.A.* **100** (7): 3960–4. doi:10.1073/pnas.0230489100. PMID 12651960. PMC 153030.
- [20] M. Ronaghi, S. Karamohamed, B. Pettersson, M. Uhlen, and P. Nyren (1996). "Real-time DNA sequencing using detection of pyrophosphate release". *Analytical Biochemistry* **242** (1): 84–9. doi:10.1006/abio.1996.0432. PMID 8923969.
- [21] US5750341 (<http://patft.uspto.gov/netacgi/nph-Parser?patentnumber=5750341>) (1995-04-17) Macevicz SC, *DNA sequencing by parallel oligonucleotide extensions*.
- [22] Hanna GJ, Johnson VA, Kuritzkes DR, *et al* (1 July 2000). "Comparison of sequencing by hybridization and cycle sequencing for genotyping of human immunodeficiency virus type 1 reverse transcriptase" (<http://jcm.asm.org/cgi/pmidlookup?view=long&pmid=10878069>). *J. Clin. Microbiol.* **38** (7): 2715–21. PMID 10878069. PMC 87006. .

- [23] J.R. Edwards, H. Ruparel, and J. Ju (2005). "Mass-spectrometry DNA sequencing". *Mutation Research* **573** (1-2): 3–12. doi:10.1016/j.mrfmmm.2004.07.021. PMID 15829234.
- [24] "VisiGen Biotechnologies Inc. - Technology Overview" (http://visigenbio.com/technology_overview.html). Visigenbio.com. . Retrieved 2009-11-15.
- [25] "The Harvard Nanopore Group" (<http://mcb.harvard.edu/branton/index.htm>). Mcb.harvard.edu. . Retrieved 2009-11-15.
- [26] "Nanopore Sequencing Could Slash DNA Analysis Costs" (<http://www.physorg.com/news157378086.html>). .
- [27] US20060029957 (<http://patft.uspto.gov/netacgi/nph-Parser?patentnumber=20060029957>) (2005-07-14) ZS Genetics, *Systems and methods of analyzing nucleic acid polymers and related components*.
- [28] "PRIZE Overview: Archon X PRIZE for Genomics" (<http://genomics.xprize.org/genomics/archon-x-prize-for-genomics/prize-overview>)
- [29] Fleischmann RD, Adams MD, White O, *et al.* (July 1995). "Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd" (<http://www.sciencemag.org/cgi/pmidlookup?view=long&pmid=7542800>). *Science* **269** (5223): 496–512. doi:10.1126/science.7542800. PMID 7542800. .
- [30] Ju J, Ruan C, Fuller CW, Glazer AN, Mathies RA (May 1995). "Fluorescence energy transfer dye-labeled primers for DNA sequencing and analysis" (<http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=41941>). *Proc. Natl. Acad. Sci. U.S.A.* **92** (10): 4347–51. doi:10.1073/pnas.92.10.4347. PMID 7753809. PMC 41941.
- [31] Ewing B, Green P (March 1998). "Base-calling of automated sequencer traces using phred. II. Error probabilities" (<http://www.genome.org/cgi/pmidlookup?view=long&pmid=9521922>). *Genome Res.* **8** (3): 186–94. doi:10.1101/gr.8.3.186 (inactive 2010-01-07). PMID 9521922. .

DNA profiling

DNA profiling (also called **DNA testing**, **DNA typing**, or **genetic fingerprinting**) is a technique employed by forensic scientists to assist in the identification of individuals on the basis of their respective DNA profiles. **DNA profiles** are encrypted sets of numbers that reflect a person's DNA makeup, which can also be used as the person's identifier. DNA profiling should not be confused with full genome sequencing.^[1] It is used in, for example, parental testing and rape investigation.

Although 99.9% of human DNA sequences are the same in every person, enough of the DNA is different to distinguish one individual from another.^[2] DNA profiling uses repetitive ("repeat") sequences that are highly variable,^[2] called variable number tandem repeats (VNTR). VNTRs loci are very similar between closely related humans, but so variable that unrelated individuals are extremely unlikely to have the same VNTRs.

The DNA profiling technique was first reported in 1984^[3] by Sir Alec Jeffreys at the University of Leicester in England,^[4] and is now the basis of several national DNA databases. Dr. Alec Jeffrey's genetic fingerprinting was made commercially available in 1987, when a chemical company, ICI, started a blood-testing center in England.^[5]

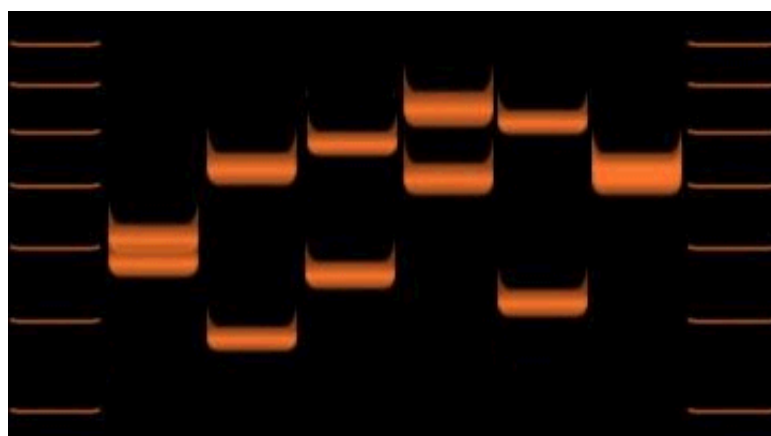
DNA profiling process

The process begins with a sample of an individual's DNA (typically called a "reference sample"). The most desirable method of collecting a reference sample is the use of a buccal swab, as this reduces the possibility of contamination. When this is not available (e.g. because a court order may be needed and not obtainable) other methods may need to be used to collect a sample of blood, saliva, semen, or other appropriate fluid or tissue from personal items (e.g. toothbrush, razor, etc) or from stored samples (e.g. banked sperm or biopsy tissue). Samples obtained from blood relatives (biological relative) can provide an indication of an individual's profile, as could human remains which had been previously profiled.

A reference sample is then analyzed to create the individual's DNA profile using one of a number of techniques, discussed below. The DNA profile is then compared against another sample to determine whether there is a genetic match.

RFLP analysis

The first methods for finding out genetics used for DNA profiling involved restriction enzyme digestion, followed by Southern blot analysis. Although polymorphisms can exist in the restriction enzyme cleavage sites, more commonly the enzymes and DNA probes were used to analyze VNTR loci. However, the Southern blot technique is laborious, and requires large amounts of undegraded sample DNA. Also, Karl Brown's original technique looked at many minisatellite loci at the same time, increasing the observed variability, but making it hard to discern individual alleles (and thereby precluding parental testing). These early techniques have been supplanted by PCR-based assays.



Variations of VNTR allele lengths in 6 individuals.

PCR analysis

With the invention of the polymerase chain reaction (PCR) technique, DNA profiling took huge strides forward in both discriminating power and the ability to recover information from very small (or degraded) starting samples. PCR greatly amplifies the amounts of a specific region of DNA, using oligonucleotide primers and a thermostable DNA polymerase. Early assays such as the HLA-DQ alpha reverse dot blot strips grew to be very popular due to their ease of use, and the speed with which a result could be obtained. However they were not as discriminating as RFLP. It was also difficult to determine a DNA profile for mixed samples, such as a vaginal swab from a sexual assault victim.

Fortunately, the PCR method is readily adaptable for analyzing VNTR loci. In the United States the FBI has standardized a set of 13 VNTR assays for DNA typing, and has organized the CODIS database for forensic identification in criminal cases. Similar assays and databases have been set up in other countries. Also, commercial kits are available that analyze single-nucleotide polymorphisms (SNPs). These kits use PCR to amplify specific regions with known variations and hybridize them to probes anchored on cards, which results in a colored spot corresponding to the particular sequence variation.

STR analysis

The method of DNA profiling used today is based on PCR and uses short tandem repeats (STR). This method uses highly polymorphic regions that have short repeated sequences of DNA (the most common is 4 bases repeated, but there are other lengths in use, including 3 and 5 bases). Because different unrelated people have different numbers of repeat units, these regions of DNA can be used to discriminate between unrelated individuals. These STR loci (locations) are targeted with sequence-specific primers and are amplified using PCR. The DNA fragments that result are then separated and detected using electrophoresis. There are two common methods of separation and detection, capillary electrophoresis (CE) and gel electrophoresis.

The polymorphisms displayed at each STR region are by themselves very common, typically each polymorphism will be shared by around 5 - 20% of individuals. When looking at multiple loci, it is the unique combination of these polymorphisms to an individual that makes this method discriminating as an identification tool. The more STR regions that are tested in an individual the more discriminating the test becomes.

From country to country, different STR-based DNA-profiling systems are in use. In North America systems which amplify the CODIS 13 core loci are almost universal, while in the UK the SGM+ system, which is compatible with The National DNA Database, is in use. Whichever system is used, many of the STR regions under test are the same. These DNA-profiling systems are based on multiplex reactions, whereby many STR regions will be under test at the same time.

The true power of STR analysis is in its statistical power of discrimination. Because the 13 loci that are currently used for discrimination in CODIS are independently assorted (having a certain number of repeats at one locus doesn't change the likelihood of having any number of repeats at any other locus), the product rule for probabilities can be applied. This means that if someone has the DNA type of ABC, where the three loci were independent, we can say that the probability of having that DNA type is the probability of having type A times the probability of having type B times the probability of having type C. This has resulted in the ability to generate match probabilities of 1 in a quintillion (1 with 18 zeros after it) or more.

However, DNA database searches showed much more frequent than expected false DNA matches including one perfect 13 locus match out of only 30,000 DNA samples in Maryland in January 2007.^[6] Moreover, since there are about 12 million monozygotic twins on Earth, that theoretical probability is useless. For example, the actual probability that 2 random persons have the same DNA depends on whether there were twins or triplets (etc.) in the family, and the number of loci used in the test. Where twins are common, the probability of matching the DNA is 22 in 1000, or about 2.2 in 100 will have matching DNA.

In practice, the risk of contaminated-matching is much greater than matching a distant relative, such as a sample being contaminated from nearby objects, or from left-over cells transferred from a prior test. Logically, the risk is greater for matching the most common person in the samples: everything collected from, or in contact with, a victim is a major source of contamination for any other samples brought into a lab. For that reason, multiple control-samples are typically tested, to ensure that they stayed clean, when prepared during the same period as the actual test samples. Unexpected matches (or variations) in several control-samples indicates a high probability of contamination for the actual test samples. In a relationship test, the full DNA profiles should differ (except for twins), to prove that a person wasn't actually matched as being related to their own DNA in another sample.

AmpFLP

Another technique, AmpFLP, or amplified fragment length polymorphism was also put into practice during the early 1990s. This technique was also faster than RFLP analysis and used PCR to amplify DNA samples. It relied on variable number tandem repeat (VNTR) polymorphisms to distinguish various alleles, which were separated on a polyacrylamide gel using an allelic ladder (as opposed to a molecular weight ladder). Bands could be visualized by silver staining the gel. One popular locus for fingerprinting was the D1S80 locus. As with all PCR based methods, highly degraded DNA or very small amounts of DNA may cause allelic dropout (causing a mistake in thinking a heterozygote is a homozygote) or other stochastic effects. In addition, because the analysis is done on a gel, very high number repeats may bunch together at the top of the gel, making it difficult to resolve. AmpFLP analysis can be highly automated, and allows for easy creation of phylogenetic trees based on comparing individual samples of DNA. Due to its relatively low cost and ease of set-up and operation, AmpFLP remains popular in lower income countries.

DNA family relationship analysis

Using PCR technology, DNA analysis is widely applied to determine genetic family relationships such as paternity, maternity, siblingship and other kinships.

During conception, the father's sperm cell and the mother's egg cell, each containing half the amount of DNA found in other body cells, meet and fuse to form a fertilized egg, called a zygote. The zygote contains a complete set of DNA molecules, a unique combination of DNA from both parents. This zygote divides and multiplies into an embryo and later, a full human being.

DNA does not change once it is formed at conception. At each stage of development, all the cells forming the body contain the same DNA—half from the father and half from the mother. This fact allows the relationship testing to use all types of all samples including loose cells from the cheeks collected using buccal swabs, blood or other types of samples.

While a lot of DNA contains information for a certain function, there is some called junk DNA, which is currently used for human identification. At some special locations (called loci) in the junk DNA, predictable inheritance patterns were found to be useful in determining biological relationships. These locations contain specific DNA markers that DNA scientists use to identify individuals. In a routine DNA paternity test, the markers used are Short Tandem Repeats (STRs), short pieces of DNA that occur in highly differential repeat patterns among individuals.

Each person's DNA contains two copies of these markers—one copy inherited from the father and one from the mother. Within a population, the markers at each person's DNA location could differ in length and sometimes sequence, depending on the markers inherited from the parents.

The combination of marker sizes found in each person makes up his/her unique genetic profile. When determining the relationship between two individuals, their genetic profiles are compared to see if they share the same inheritance patterns at a statistically conclusive rate.

For example, the following sample report from this commercial DNA paternity testing laboratory Universal Genetics [7] signifies how relatedness between parents and child is identified on those special markers:

DNA Marker	Mother	Child	Alleged Father
D21S11	28, 30	28, 31	29, 31
D7S820	9, 10	10, 11	11, 12
TH01	14, 15	14, 16	15, 16
D13S317	7, 8	7, 9	8, 9
D19S433	14, 16.2	14, 15	15, 17

The partial results indicate that the child and the alleged father's DNA match among these five markers. The complete test results show this correlation on 16 markers between the child and the tested man to draw a conclusion of whether or not the man is the biological father.

Scientifically, each marker is assigned with a Paternity Index (PI), which is a statistical measure of how powerfully a match at a particular marker indicates paternity. The PI of each marker is multiplied with each other to generate the Combined Paternity Index (CPI), which indicates the overall probability of an individual being the biological father of the tested child relative to any random man from the entire population of the same race. The CPI is then converted into a Probability of Paternity showing the degree of relatedness between the alleged father and child.

The DNA test report in other family relationship tests, such as grandparentage and siblingship tests, is similar to a paternity test report. Instead of the Combined Paternity Index, a different value, such as a Siblingship Index, is reported.

The report shows the genetic profiles of each tested person. If there are markers shared among the tested individuals, the probability of biological relationship is calculated to determine how likely the tested individuals share the same

markers due to a blood relationship.

Y-chromosome analysis

Recent innovations have included the creation of primers targeting polymorphic regions on the Y-chromosome (Y-STR), which allows resolution of a mixed DNA sample from a male and female and/or cases in which a differential extraction is not possible. Y-chromosomes are paternally inherited, so Y-STR analysis can help in the identification of paternally related males. Y-STR analysis was performed in the Sally Hemings controversy to determine if Thomas Jefferson had sired a son with one of his slaves.

Mitochondrial analysis

For highly degraded samples, it is sometimes impossible to get a complete profile of the 13 CODIS STRs. In these situations, mitochondrial DNA (mtDNA) is sometimes typed due to there being many copies of mtDNA in a cell, while there may only be 1-2 copies of the nuclear DNA. Forensic scientists amplify the HV1 and HV2 regions of the mtDNA, then sequence each region and compare single-nucleotide differences to a reference. Because mtDNA is maternally inherited, directly linked maternal relatives can be used as match references, such as one's maternal grandmother's daughter's son. A difference of two or more nucleotides is generally considered to be an exclusion. Heteroplasmy and poly-C differences may throw off straight sequence comparisons, so some expertise on the part of the analyst is required. mtDNA is useful in determining clear identities, such as those of missing persons when a maternally linked relative can be found. mtDNA testing was used in determining that Anna Anderson was not the Russian princess she had claimed to be, Anastasia Romanov.

mtDNA can be obtained from such material as hair shafts and old bones/teeth..

DNA databases

There are now several DNA databases in existence around the world. Some are private, but most of the largest databases are government controlled. The United States maintains the largest DNA database, with the Combined DNA Index System, holding over 5 million records as of 2007.^[8] The United Kingdom maintains the National DNA Database (NDNAD), which is of similar size, despite the UK's smaller population. The size of this database, and its rate of growth, is giving concern to civil liberties groups in the UK, where police have wide-ranging powers to take samples and retain them even in the event of acquittal.^[9]

The U.S. Patriot Act of the United States provides a means for the U.S. government to get DNA samples from other countries if they are either a division of, or head office of, a company operating in the U.S. Under the act, the American offices of the company can't divulge to their subsidiaries/offices in other countries the reasons that these DNA samples are sought or by whom.

When a match is made from a National DNA Databank to link a crime scene to an offender who has provided a DNA Sample to a databank that link is often referred to as a *cold hit*. A cold hit is of value in referring the police agency to a specific suspect but is of less evidential value than a DNA match made from outside the DNA Databank.^[10]

Considerations when evaluating DNA evidence

In the early days of the use of genetic fingerprinting as criminal evidence, juries were often swayed by spurious statistical arguments by defense lawyers along these lines: given a match that had a 1 in 5 million probability of occurring by chance, the lawyer would argue that this meant that in a country of say 60 million people there were 12 people who would also match the profile. This was then translated to a 1 in 12 chance of the suspect being the guilty one. This argument is not sound unless the suspect was drawn at random from the population of the country. In fact, a jury should consider how likely it is that an individual matching the genetic profile would also have been a suspect

in the case for other reasons. Another spurious statistical argument is based on the false assumption that a 1 in 5 million probability of a match automatically translates into a 1 in 5 million probability of innocence and is known as the prosecutor's fallacy.

When using RFLP, the theoretical risk of a coincidental match is 1 in 100 billion (100,000,000,000), although the practical risk is actually 1 in 1000 because monozygotic twins are 0.2% of the human population. Moreover, the rate of laboratory error is almost certainly higher than this, and often actual laboratory procedures do not reflect the theory under which the coincidence probabilities were computed. For example, the coincidence probabilities may be calculated based on the probabilities that markers in two samples have bands in *precisely* the same location, but a laboratory worker may conclude that similar—but not precisely identical—band patterns result from identical genetic samples with some imperfection in the agarose gel. However, in this case, the laboratory worker increases the coincidence risk by expanding the criteria for declaring a match. Recent studies have quoted relatively high error rates which may be cause for concern.^[11] In the early days of genetic fingerprinting, the necessary population data to accurately compute a match probability was sometimes unavailable. Between 1992 and 1996, arbitrary low ceilings were controversially put on match probabilities used in RFLP analysis rather than the higher theoretically computed ones.^[12] Today, RFLP has become widely disused due to the advent of more discriminating, sensitive and easier technologies.

STRs do not suffer from such subjectivity and provide similar power of discrimination (1 in 10^{13} for unrelated individuals if using a full SGM+ profile) It should be noted that figures of this magnitude are not considered to be statistically supportable by scientists in the UK, for unrelated individuals with full matching DNA profiles a match probability of 1 in a billion is considered statistically supportable (Since 1998 the DNA profiling system supported by The National DNA Database in the UK is the SGM+ DNA profiling system which includes 10 STR regions and a sex indicating test. However, with any DNA technique, the cautious juror should not convict on genetic fingerprint evidence alone if other factors raise doubt. Contamination with other evidence (secondary transfer) is a key source of incorrect DNA profiles and raising doubts as to whether a sample has been adulterated is a favorite defense technique. More rarely, Chimerism is one such instance where the lack of a genetic match may unfairly exclude a suspect.

Evidence of genetic relationship

It's also possible to use DNA profiling as evidence of genetic relationship, but testing that shows no relationship isn't absolutely certain. While almost all individuals have a single and distinct set of genes, rare individuals, known as "chimeras", have at least two different sets of genes. There have been several cases of DNA profiling that falsely "proved" that a mother was unrelated to her children.^[13]

Fake DNA evidence

The value of DNA evidence has to be seen in light of recent cases where criminals planted fake DNA samples at crime scenes. In one case, a criminal even planted fake DNA evidence in his own body: Dr. John Schneeberger raped one of his sedated patients in 1992 and left semen on her underwear. Police drew what they believed to be Schneeberger's blood and compared its DNA against the crime scene semen DNA on three occasions, never showing a match. It turned out that he had surgically inserted a Penrose drain into his arm and filled it with foreign blood and anticoagulants.

In a study conducted by the life science company Nucleix and published in the journal Forensic Science International, scientists found that an In vitro synthesized sample of DNA matching any desired genetic profile can be constructed using standard molecular biology techniques without obtaining any actual tissue from that person.

DNA evidence as evidence in criminal trials

Familial searching

Familial searching is the use of family members' DNA to identify a closely related suspect in jurisdictions where large DNA databases exist, but no exact match has been found. The first successful use of the practice was in a UK case where a man was convicted of manslaughter when he threw a brick stained with his own blood into a moving car. Police could not get an exact match to the UK's DNA database because the man had no criminal convictions, but police implicated him using a close relative's DNA.^[14]

Surreptitious DNA collecting

Police forces may collect DNA samples without the suspects' knowledge, and use it as evidence. Legality of this mode of proceeding has been questioned in Australia.

In the United States, it has been accepted, courts often claiming that there was no expectation of privacy, citing *California v. Greenwood* (1985), during which the Supreme Court held that the Fourth Amendment does not prohibit the warrantless search and seizure of garbage left for collection outside the curtilage of a home. Critics of this practice underline the fact that this analogy ignores that "most people have no idea that they risk surrendering their genetic identity to the police by, for instance, failing to destroy a used coffee cup. Moreover, even if they do realize it, there is no way to avoid abandoning one's DNA in public."^[15]

In the UK, the Human Tissue Act of 2004 prohibited private individuals from covertly collecting biological samples (hair, fingernails, etc.) for DNA analysis, but excluded medical and criminal investigations from the offense.^[16]

England and Wales

Evidence from an expert who has compared DNA samples must be accompanied by evidence as to the sources of the samples and the procedures for obtaining the DNA profiles.^[17] The judge must ensure that the jury must understand the significance of DNA matches and mismatches in the profiles. The judge must also ensure that the jury does not confuse the 'match probability' (the probability that a person that is chosen at random has a matching DNA profile to the sample from the scene) with the 'likelihood ratio' (the probability that a person with matching DNA committed the crime). In *R v. Doheny*, EWCA Crim 728^[18]. Phillips LJ gave this example of a summing up, which should be carefully tailored to the particular facts in each case:

Members of the Jury, if you accept the scientific evidence called by the Crown, this indicates that there are probably only four or five white males in the United Kingdom from whom that semen stain could have come. The Defendant is one of them. If that is the position, the decision you have to reach, on all the evidence, is whether you are sure that it was the Defendant who left that stain or whether it is possible that it was one of that other small group of men who share the same DNA characteristics.

Juries should weigh up conflicting and corroborative evidence, using their own common sense and not by using mathematical formulae, such as Bayes' theorem, so as to avoid "confusion, misunderstanding and misjudgment".^[19]

Presentation and evaluation of evidence of partial or incomplete DNA profiles

R v Bates (2006) EWCA Crim 1395 Moore-Bick LJ said:

"We can see no reason why partial profile DNA evidence should not be admissible provided that the jury are made aware of its inherent limitations and are given a sufficient explanation to enable them to evaluate it. There may be cases where the match probability in relation to all the samples tested is so great that the judge would consider its probative value to be minimal and decide to exclude the evidence in the exercise of his discretion, but this gives rise to no new question of principle and can be left for decision on a case by case basis. However, the fact that there exists in the case of all partial profile evidence the possibility that a

"missing" allele might exculpate the accused altogether does not provide sufficient grounds for rejecting such evidence. In many there is a possibility (at least in theory) that evidence exists which would assist the accused and perhaps even exculpate him altogether, but that does not provide grounds for excluding relevant evidence that is available and otherwise admissible, though it does make it important to ensure that the jury are given sufficient information to enable them to evaluate that evidence properly".^[20]

DNA testing in the US

There are state laws on DNA profiling in all 50 states of the United States.^[21] Detailed information on database laws in each state can be found at the National Conference of State Legislatures website.^[22]

Development of artificial DNA

In August 2009, scientists in Israel stunned the forensic sciences and raised serious questions concerning law enforcement's use of DNA matching as the ultimate method of identification. In a paper published in the journal *Forensic Science International: Genetics*, the Israeli researchers demonstrated that it is possible to manufacture DNA in a laboratory, and thus falsify DNA evidence. The scientists had fabricated saliva and blood samples, which originally contained DNA from a person other than the ostensible donor of the blood and saliva.^[23]

Additionally, and perhaps more frighteningly, the same researchers showed that, using a DNA database, it is possible to take information from a profile and actually manufacture DNA to match it. Worse, this can be done without access to any actual DNA from the person whose DNA they are duplicating. The synthetic DNA oligos required for the procedure are used in probably every molecular laboratory.^[23]

Dr. Daniel Frumkin, lead author on the paper, was quoted in *The New York Times* as saying, "You can just engineer a crime scene...any biology undergraduate could perform this."^[23]

Dr. Frumkin fortunately has perfected a test that can forensically differentiate real DNA samples from fake ones. His test uses epigenetic modifications, in particular, DNA methylation. Seventy percent of the DNA in any human genome is methylated, meaning it contains methyl group modifications within a CpG dinucleotide context. Methylation at the promoter region is associated with gene silencing. It appears that the synthetic DNA lacks this epigenetic modification, which allows the test to be used to distinguish manufactured DNA from original, genuine, DNA.^[23]

The idea that DNA can be fabricated, and then planted at a crime scene, is now reality; fortunately Dr. Frumkin has developed the test to differentiate real from fake DNA. But it is unknown how many, if any, police departments currently use the test, which is distressing considering Frumkin's claim that the DNA manufacturing procedure is within the grasp of any undergraduate biology student. No police lab has publicly announced that it is using the new test to verify DNA results, while FSI Genetics says that any forensic laboratory doing DNA identification should adopt this test to authenticate its results as "real" DNA.^[24]

Cases

- In the 1950s, Anna Anderson claimed that she was Grand Duchess Anastasia Nikolaevna of Russia; in the 1980s after her death, samples of her tissue that had been stored at a Charlottesville, Virginia hospital following a medical procedure were tested using DNA fingerprinting and showed that she bore no relation to the Romanovs.^[25]
- In 1986, Richard Buckland was exonerated despite having admitted to the rape and murder of a teenager near Leicester, the city where DNA profiling was first discovered. This was the first use of DNA finger printing in a criminal investigation.^[26]
- In 1987 it was the first time genetic fingerprinting was used in criminal court where a man accused of unlawful intercourse with a mentally handicapped 14-year-old female who gave birth to his baby.^[27]

- In 1987, in the same case as Buckland, British baker Colin Pitchfork was the first criminal caught and convicted using DNA fingerprinting.^[28]
 - In 1987, Florida rapist Tommy Lee Andrews was the first person in the United States to be convicted as a result of DNA evidence, for raping a woman during a burglary; he was convicted on 6 November 1987 and sentenced to 22 years in prison.^{[29] [30]}
 - In 1988, Timothy Spencer was the first man in Virginia to be sentenced to death through DNA testing, for several rape and murder charges; he was dubbed "The South Side Strangler" because he killed victims on the southside of Richmond, Virginia. He was later charged with rape and 1st degree murder and was sentenced to death. He was executed on April 27, 1994. David Vasquez, initially convicted of one of Spencer's crimes, became the first man in America exonerated based on DNA evidence.
 - In 1989, Chicago man Gary Dotson was the first person whose conviction was overturned using DNA evidence.
 - In 1991, Allan Legere was the first Canadian to be convicted as a result of DNA evidence, for four murders he had committed while an escaped prisoner in 1989. During his trial, his defense argued that the relatively shallow gene pool of the region could lead to false positives.
 - In 1992, DNA evidence was used to prove that Nazi doctor Josef Mengele was buried in Brazil under the name Wolfgang Gerhard.
 - In 1993, Kirk Bloodsworth was the first person to have been convicted of murder and sentenced to death, whose conviction was overturned using DNA evidence.
 - The 1993 rape and murder of Mia Zapata, lead singer for the Seattle punk band The Gits was unsolved 9 years after the murder. A database search in 2001 failed, but the killer's DNA was collected when he was arrested in Florida for burglary and domestic abuse in 2002.
 - The science was made famous in the United States in 1994 when prosecutors heavily relied on — and through expert witnesses exhaustively presented and explained — DNA evidence allegedly linking O.J. Simpson to a double murder. The case also brought to light the laboratory difficulties and handling procedure mishaps which can cause such evidence to be significantly doubted.
 - In 1994, Royal Canadian Mounted Police (RCMP) detectives successfully tested hairs from a cat known as Snowball, and used the test to link a man to the murder of his wife, thus marking for the first time in forensic history the use of non-human DNA to identify a criminal.
 - In 1998, Dr. Richard J. Schmidt was convicted of attempted second-degree murder when it was shown that there was a link between the viral DNA of the human immunodeficiency virus (HIV) he had been accused of injecting in his girlfriend and viral DNA from one of his patients with full-blown AIDS. This was the first time viral DNA fingerprinting had been used as evidence in a criminal trial.
 - In 1999, Raymond Easton a disabled man from Swindon, England was arrested and detained for 7 hours in connection with a burglary due to an inaccurate DNA match. His DNA had been retained on file after an unrelated domestic incident some time previously.^[31]
 - In May 2000 Gordon Graham murdered Paul Gault at his home in Lisburn, Northern Ireland. Graham was convicted of the murder when his DNA was found on a sports bag left in the house as part of an elaborate ploy to suggest the murder occurred after a burglary had gone wrong. Graham was having an affair with the victims wife at the time of the murder. It was the first time Low Copy Number DNA was used in Northern Ireland. Media:<http://www.belfasttelegraph.co.uk/sunday-life/news/freedom-in-bag-for-killer-graham-13906320.html>
- In 2001, Wayne Butler was convicted for the murder of Celia Douty. It was the first murder in Australia to be solved using DNA profiling.^{[32] [33]}
- In 2002, DNA testing was used to exonerate Douglas Echols, a man who was wrongfully convicted in a 1986 rape case. Echols was the 114th person to be exonerated through post-conviction DNA testing.
 - In August 2002, Annalisa Vincenzi was shot dead in Tuscany. Some time later, Bartender Peter Hamkin, 23, was arrested in Merseyside in March 2003 on an extradition warrant heard at Bow Street Magistrates' Court in London to establish whether he should be taken to Italy to face a murder charge. DNA "proved" he shot her, but he was

cleared on other evidence.^[34]

- In 2003, Welshman Jeffrey Gafoor was convicted of the 1988 murder of Lynette White, when crime scene evidence collected 12 years earlier was re-examined using STR techniques, resulting in a match with his nephew.^[35] This may be the first known example of the DNA of an innocent yet related individual being used to identify the actual criminal, via "familial searching".
- In June 2003, because of new DNA evidence, Dennis Halstead, John Kogut and John Restivo won a re-trial on their murder conviction. The three men had already served eighteen years of their thirty-plus-year sentences.
- The trial of Robert Pickton is notable in that DNA evidence is being used primarily to identify the *victims*, and in many cases to prove their existence.
- In March 2003, Josiah Sutton was released from prison after serving four years of a twelve-year sentence for a sexual assault charge. Questionable DNA samples taken from Sutton were retested in the wake of the Houston Police Department's crime lab scandal of mishandling DNA evidence.
- In 2004, DNA testing shed new light into the mysterious 1912 disappearance of Bobby Dunbar, a four-year-old boy who vanished during a fishing trip. He was allegedly found alive eight months later in the custody of William Cantwell Walters, but another woman claimed that the boy was her son, Bruce Anderson, whom she had entrusted in Walters' custody. The courts disbelieved her claim and convicted Walters for the kidnapping. The boy was raised and known as Bobby Dunbar throughout the rest of his life. However, DNA tests on Dunbar's son and nephew revealed the two were not related, thus establishing that the boy found in 1912 was not Bobby Dunbar, whose real fate remains unknown.^[36]
- In 2005, Gary Leiterman was convicted of the 1969 murder of Jane Mixer, a law student at the University of Michigan, after DNA found on Mixer's pantyhose was matched to Leiterman. DNA in a drop of blood on Mixer's hand was matched to John Ruelas, who was only four years old in 1969 and was never successfully connected to the case in any other way. Leiterman's defense unsuccessfully argued that the unexplained match of the blood spot to Ruelas pointed to cross-contamination and raised doubts about the reliability of the lab's identification of Leiterman.^[37] ^[38] ^[39]
- In December 2005, Evan Simmons was proven innocent of a 1981 attack on an Atlanta woman after serving twenty-four years in prison. Mr Clark is the 164th person in the United States and the fifth in Georgia to be freed using post-conviction DNA testing.
- In March 2009, Sean Hodgson who spent 27 years in jail, convicted of killing Teresa De Simone, 22, in her car in Southampton 30 years ago was quashed by senior judges. Tests prove DNA from the scene was not his. British police have now reopened the case.

See also

- capillary electrophoresis (CE)
 - Forensic identification
 - Full genome sequencing
 - Gene mapping
 - genealogical DNA test
 - *Harvey v. Horan*
 - Identification (biology)
 - National DNA database
 - Parental testing
 - Phantom of Heilbronn
 - Project Innocence
 - restriction fragment length polymorphism (RFLP)
 - ribotyping
 - short tandem repeat (STR)
-

- State of Louisiana v. Frisard

External links

- [40] Eureka moment that led to the discovery of DNA fingerprinting
- EHSTRAFD.org - Earth Human STR Allele Frequencies Database ^[41]
- Why A DNA Test should include the Mothers DNA ^[42]
- How to Choose a Qualified DNA Laboratory ^[43]
- HaploGroups.com ^[44] The most comprehensive resource for DNA testing to help you find your genetic origin or Haplogroup
- How to make a DNA (genetic) Fingerprint and applications of DNA Fingerprinting ^[45]
- Create a DNA Fingerprint ^[46]
- Fingerprinting.com ^[47] DNA Fingerprinting Identification and Methods
- Forensic genetics and ethical, legal and social implications beyond the clinic ^[48]
- In silico simulation of Molecular Biology Techniques ^[49] - A place to learn typing techniques by simulating them
- The History of DNA Testing ^[50] Interactive presentation uncovering the historical origins of DNA testing and highlighting famous paternity testing cases. (requires Adobe Flash)
- [51] National DNA-Databases in the EU (Nathan Van Camp & Kris Dierickx)
- NPIA NDNAD Website ^[52]
- The Evaluation of Forensic DNA Evidence ^[53]
- Key Dates in the History of DNA Profiling ^[54]

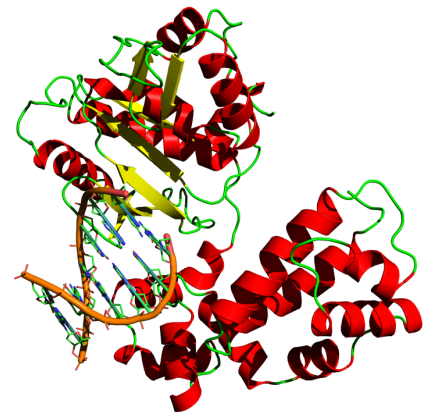
References

- [1] Kijk magazine, 01 January 2009
- [2] "Use of DNA in Identification" (http://www.accessexcellence.org/RC/AB/BA/Use_of_DNA_Identification.php). Accessexcellence.org. . Retrieved 2010-04-03.
- [3] Joseph Wambaugh, *The Blooding* (New York, New York: A Perigord Press Book, 1989), 83.
- [4] Jeffreys A.J., Wilson V., Thein S.W. (1984). "Hypervariable 'minisatellite' regions in human DNA" (<http://www.nature.com/nature/journal/v314/n6006/abs/314067a0.html>). *Nature* 314: 67–73. doi:10.1038/314067a0
- [5] Joseph Wambaugh, *The Blooding* (New York, New York: A Perigord Press Book, 1989), 202.
- [6] Felch, Jason; et al (July 20, 2008). "FBI resists scrutiny of 'matches'" (<http://articles.latimes.com/2008/jul/20/local/me-dna20>). *Los Angeles Times*: pp. P8. .
- [7] "Leading DNA testing lab in Los Angeles providing paternity test and forensic DNA test" (<http://www.dnatestingforpaternity.com>). Dnatestingforpaternity.com. . Retrieved 2010-04-03.
- [8] "CODIS - National DNA Index System" (<http://www.fbi.gov/hq/lab/codis/national.htm>). Fbi.gov. . Retrieved 2010-04-03.
- [9] "Restrictions on use and destruction of fingerprints and samples" (http://www.wikicrimeline.co.uk/index.php?title=Identification_by_body_samples_and_impressions#Section_82:_Restrictions_on_use_and_destruction_of_fingerprints_and_samples). Wikicrimeline.co.uk. 2009-09-01. . Retrieved 2010-04-03.
- [10] Rose & Goos "DNA - A Practical Guide" (Carswell Publications, Toronto)
- [11] Nick Paton Walsh False result fear over DNA tests (<http://www.guardian.co.uk/crime/article/0,2763,640157,00.html>) *The Observer*, Sunday 27 January 2002
- [12] The Evaluation of Forensic DNA Evidence (<http://www.nap.edu/openbook/0309053951/html/35.html>) 1996
- [13] "Two Women Don't Match Their Kids' DNA" (<http://abcnews.go.com/Primetime/story?id=2315693>). Abcnews.go.com. 2006-08-15. . Retrieved 2010-04-03.
- [14] Bhattacharya, Shaoni (20 April 2004). "Killer convicted thanks to relative's DNA" (<http://www.newscientist.com/article/dn4908-killer-convicted-thanks-to-relatives-dna.html>). *New Scientist*. . Retrieved 2008-05-06.
- [15] Amy Harmon, Lawyers Fight DNA Samples Gained on Sly (<http://www.nytimes.com/2008/04/03/science/03dna.html>), *New York Times*, April 3, 2008.
- [16] Human Tissue Act 2004 (http://www.dh.gov.uk/en/PublicHealth/Scientificdevelopmentgeneticsandbioethics/Tissue/TissueGeneralInformation/DH_4102169), UK, available in pdf.
- [17] *R v. Loveridge*, EWCA Crim 734 .
- [18] <http://www.bailii.org/cgi-bin/markup.cgi?doc=/ew/cases/EWCA/Crim/1996/728.html&query=doheny&method=boolean>

- [19] *R v. Adams*, EWCA Crim 2474 (<http://www.bailii.org/cgi-bin/markup.cgi?doc=/ew/cases/EWCA/Crim/1997/2474.html&query=doheny&method=boolean>).
- [20] "WikiCrimeLine DNA profiling" (http://www.wikicrimeline.co.uk/index.php?title=DNA_profiling#Presentation_and_evaluation_of_evidence_of_partial_or_incomplete_DNA_profiles). Wikicrimeline.co.uk. Retrieved 2010-04-03.
- [21] "Genelex: The DNA Paternity Testing Site" (<http://www.healthanddna.com/dna-learning/book-dna-testing.html>). Healthanddna.com. 1996-01-06. Retrieved 2010-04-03.
- [22] Donna Lyons - Posted by Glenda. "State Laws on DNA Data Banks" (<http://www.ncsl.org/IssuesResearch/CivilandCriminalJustice/StateLawsonDNADatabanks/tabid/12737/Default.aspx>). Ncsl.org. Retrieved 2010-04-03.
- [23] Pollack, Andrew (August 18, 2009). "DNA Evidence Can Be Fabricated, Scientists Show" (http://www.nytimes.com/2009/08/18/science/18dna.html?_r=1). *The New York Times*. Retrieved April 1, 2010.
- [24] "Elsevier" ([http://www.fsigenetics.com/article/S1872-4973\(09\)00099-4/abstract](http://www.fsigenetics.com/article/S1872-4973(09)00099-4/abstract)). Fsigenetics.com. Retrieved 2010-04-03.
- [25] Identification of the remains of the Romanov family by DNA analysis by Peter Gill, Central Research and Support Establishment, Forensic Science Service, Aldermaston, Reading, Berkshire, RG7 4PN, UK, Pavel L. Ivanov, Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, 117984, Moscow, Russia, Colin Kimpton, Romelle Piercy, Nicola Benson, Gillian Tully, Ian Evett, Kevin Sullivan, Forensic Science Service, Priory House, Gooch Street North, Birmingham B5 6QQ, UK, Erika Hagelberg, University of Cambridge, Department of Biological Anthropology, Downing Street, Cambridge CB2 3DZ, UK - (<http://www.nature.com/ng/journal/v6/n2/abs/ng0294-130.html>)
- [26] "DNA pioneer's 'eureka' moment" (<http://news.bbc.co.uk/1/hi/programmes/newsnight/8245312.stm>). *BBC News*. September 9, 2009. Retrieved April 1, 2010.
- [27] Joseph Wambaugh, *The Blooding* (New York, New York: A Perigord Press Book, 1989), 316.
- [28] Joseph Wambaugh, *The Blooding* (New York, New York: A Perigord Press Book, 1989), 369.
- [29] "Gene Technology - Page 14" (<http://www.txtwriter.com/Backgrounders/Genetech/GEpage14.html>). Txtwriter.com. 1987-11-06. Retrieved 2010-04-03.
- [30] "frontline: the case for innocence: the dna revolution: state and federal dna database laws examined" (<http://www.pbs.org/wgbh/pages/frontline/shows/case/revolution/databases.html>). Pbs.org. Retrieved 2010-04-03.
- [31] Jeffries, Stuart (2006-10-08). "Suspect Nation" (<http://www.guardian.co.uk/comment/story/0,,1933724,00.html>). London: The Guardian. Retrieved April 1, 2010.
- [32] Dutter, Barbie (2001-06-19). "18 years on, man is jailed for murder of Briton in 'paradise'" (<http://www.telegraph.co.uk/news/worldnews/australiaandthepacific/1322624/18-years-on,-man-is-jailed-for-murder-of-Briton-in-'paradise'.html>). London: The Telegraph. Retrieved 2008-06-17.
- [33] McCutcheon, Peter (2004-09-08). "DNA evidence may not be infallible: experts" (<http://www.abc.net.au/7.30/content/2004/s1195029.htm>). Australian Broadcasting Corporation. Retrieved 2008-06-17.
- [34] "Mistaken identity claim over murder" (<http://news.bbc.co.uk/1/hi/england/2766289.stm>). *BBC News*. February 15, 2003. Retrieved April 1, 2010.
- [35] Satish Sekar. "Lynette White Case: How Forensics Caught the Cellophane Man" (<http://lifeloom.com/l2Sekar.htm>). Lifeloom.com. Retrieved 2010-04-03.
- [36] "DNA clears man of 1914 kidnapping conviction (http://www.usatoday.com/tech/news/2004-05-05-1914-dna_x.htm)," *USA Today*, (May 5, 2004), by Allen G. Breed, Associated Press.
- [37] CBS News (<http://www.cbsnews.com/stories/2005/11/22/48hours/main1066064.shtml>) story on the Jane Mixer murder case; March 24, 2007.
- [38] Another CBS News (<http://www.cbsnews.com/stories/2007/07/17/earlyshow/main3065082.shtml>) story on the Mixer case; July 17, 2007.
- [39] An advocacy site (<http://www.garyisinnocent.org>) challenging Leiterman's conviction in the Mixer murder.
- [40] <http://www.guardian.co.uk/science/2009/may/24/dna-fingerprinting-alec-jeffreys>
- [41] <http://www.ehstrafd.org>
- [42] <http://www.dna-worldwide.com/resource-centre/paternity-test-info/why-test-mum-s-dna/>
- [43] <http://www.dnatestingforpaternity.com/choose-dna-testing-lab.html>
- [44] <http://www.HaploGroups.com>
- [45] <http://www.thenakedscientists.com/html/columnists/dalyacolumn8.htm>
- [46] <http://www.pbs.org/wgbh/nova/sheppard/analyze.html>
- [47] <http://www.fingerprinting.com/dna-fingerprinting.php>
- [48] <http://www.nature.com/ng/journal/v36/n11s/abs/ng1594.html>
- [49] <http://insilico.ehu.es>
- [50] <http://www.dnacenter.com/science-technology/articles/history-of-dna-testing.html>
- [51] <https://www.kuleuven.be/cbmer/page.php?LAN=E&ID=399&FILE=subject&PAGE=1>
- [52] <http://www.npia.police.uk/en/8934.htm>
- [53] http://www.nap.edu/catalog.php?record_id=5141
- [54] http://www.crimtrac.gov.au/systems_projects/KeyDatesintheHistoryofDNAProfiling.html

DNA polymerase

A **DNA polymerase** is an enzyme that catalyzes the polymerization of deoxyribonucleotides into a DNA strand. DNA polymerases are best-known for their role in DNA replication, in which the polymerase "reads" an intact DNA strand as a template and uses it to synthesize the new strand. This process copies a piece of DNA. The newly-polymerized molecule is complementary to the template strand and identical to the template's original partner strand. DNA polymerases use a magnesium ion for catalytic activity.



3D structure of the DNA-binding helix-turn-helix motifs in human DNA polymerase beta

Function

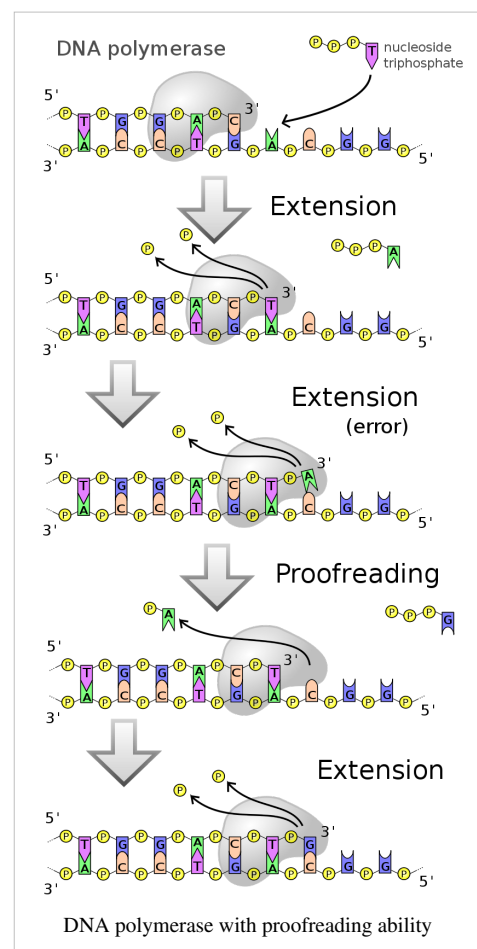
DNA polymerase can add free nucleotides to only the 3' end of the newly-forming strand. This results in elongation of the new strand in a 5'-3' direction. No known DNA polymerase is able to begin a new chain (*de novo*). DNA polymerase can add a nucleotide onto only a preexisting 3'-OH group, and, therefore, needs a primer at which it can add the first nucleotide. Primers consist of RNA and DNA bases with the first two bases always being RNA, and are synthesized by another enzyme called primase. An enzyme known as a helicase is required to unwind DNA from a double-strand structure to a single-strand structure to facilitate replication of each strand consistent with the semiconservative model of DNA replication.

Error correction is a property of some, but not all, DNA polymerases. This process corrects mistakes in newly-synthesized DNA. When an incorrect base pair is recognized, DNA polymerase reverses its direction by one base pair of DNA. The 3'-5' exonuclease activity of the enzyme allows the incorrect base pair to be excised (this activity is known as *proofreading*). Following base excision, the polymerase can re-insert the correct base and replication can continue.

Various DNA polymerases are extensively used in molecular biology experiments.

Variation across species

DNA polymerases have highly-conserved structure, which means that their overall catalytic subunits vary, on a whole, very little from species to species. Conserved structures usually indicate important, irreplicable functions of the cell, the maintenance of which provides evolutionary advantages.



Some viruses also encode special DNA polymerases, such as Hepatitis B virus DNA polymerase. These may selectively replicate viral DNA through a variety of mechanisms. Retroviruses encode an unusual DNA polymerase called reverse transcriptase, which is an RNA-dependent DNA polymerase (RdDp). It polymerizes DNA from a template of RNA.

DNA polymerase families

Based on sequence homology, DNA polymerases can be further subdivided into seven different families: A, B, C, D, X, Y, and RT.

Family A

Polymerases contain both replicative and repair polymerases. Replicative members from this family include the extensively-studied T7 DNA polymerase, as well as the eukaryotic mitochondrial DNA Polymerase γ . Among the repair polymerases are *Escherichia coli* DNA pol I, *Thermus aquaticus* pol I, and *Bacillus stearothermophilus* pol I. These repair polymerases are involved in excision repair and processing of Okazaki fragments generated during lagging strand synthesis.

Family B

Polymerases mostly contain replicative polymerases and include the major eukaryotic DNA polymerases α , δ , ϵ , (see Greek letters) and also DNA polymerase ζ . Family B also includes DNA polymerases encoded by some bacteria and bacteriophages, of which the best-characterized are from T4, Phi29, and RB69 bacteriophages. These enzymes are involved in both leading and lagging strand synthesis during replication. A hallmark of the B family of polymerases is their highly faithful DNA synthesis during replication. While many have an intrinsic 3'-5' proofreading exonuclease activity, eukaryotic DNA polymerases α and ζ are two examples of B family polymerases lacking this proofreading activity.

Family C

Polymerases are the primary bacterial chromosomal replicative enzymes. DNA Polymerase III alpha subunit from *E. coli* is the catalytic subunit ^[1] and possesses no known nuclease activity. A separate subunit, the epsilon subunit, possesses the 3'-5' exonuclease activity used for editing during chromosomal replication. Recent research has classified Family C polymerases as a subcategory of Family X.

Family D

Polymerases are still not very well characterized. All known examples are found in the Euryarchaeota subdomain of Archaea and are thought to be replicative polymerases.

Family X

Contains the well-known eukaryotic polymerase pol β , as well as other eukaryotic polymerases such as pol σ , pol λ , pol μ , and terminal deoxynucleotidyl transferase (TdT). Pol β is required for short-patch base excision repair, a DNA repair pathway that is essential for repairing abasic sites. Pol λ and Pol μ are involved in non-homologous end-joining, a mechanism for rejoining DNA double-strand breaks. TdT is expressed only in lymphoid tissue, and adds "n nucleotides" to double-strand breaks formed during V(D)J recombination to promote immunological diversity. The yeast *Saccharomyces cerevisiae* has only one Pol X polymerase, Pol4, which is involved in non-homologous end-joining.

Family Y

Polymerases differ from others in having a low fidelity on undamaged templates and in their ability to replicate through damaged DNA. Members of this family are hence called translesion synthesis (TLS) polymerases. Depending on the lesion, TLS polymerases can bypass the damage in an error-free or error-prone fashion, the latter resulting in elevated mutagenesis. Xeroderma pigmentosum variant (XPV) patients for instance have mutations in the gene encoding Pol η (eta), which is error-free for UV-lesions. In XPV patients, alternative error-prone polymerases, e.g., Pol ζ (zeta) (polymerase ζ is a B Family polymerase a complex of the catalytic subunit REV3L with Rev7, which associates with Rev1^[2]), are thought to be involved in mistakes that result in the cancer predisposition of these patients. Other members in humans are Pol ι (iota), Pol κ (kappa), and Rev1 (terminal deoxycytidyl transferase). In *E. coli*, two TLS polymerases, Pol IV (DINB) and PolV (UmuD'₂C), are known.

Family RT

The reverse transcriptase family contains examples from both retroviruses and eukaryotic polymerases. The eukaryotic polymerases are usually restricted to telomerases. These polymerases use an RNA template to synthesize the DNA strand.

Prokaryotic DNA polymerases

Bacteria have 5 known DNA polymerases:

- **Pol I:** implicated in DNA repair; has 5'→3' (Polymerase) activity and both 3'→5' exonuclease (Proofreading) and 5'→3' exonuclease activity (RNA Primer removal).
- **Pol II:** involved in reparation of damaged DNA; has 3'→5' exonuclease activity.
- **Pol III:** the main polymerase in bacteria (elongates in DNA replication); has 3'→5' exonuclease proofreading ability.
- **Pol IV:** a Y-family DNA polymerase.
- **Pol V:** a Y-family DNA polymerase; participates in bypassing DNA damage.

Eukaryotic DNA polymerases

Eukaryotes have at least 15 DNA Polymerases:^[3]

- **Pol α** (synonyms are RNA primase, **DNA** polymerase): forms a complex with a small catalytic (PriS) and a large noncatalytic (PriL) subunit^[4], with the Pri subunits acting as a primase (synthesizing an RNA primer), and then with DNA Pol α elongating that primer with DNA nucleotides. After around 20 nucleotides^[5] elongation is taken over by Pol ϵ (on the leading strand) and δ (on the lagging strand).
- **Pol β :** Implicated in repairing DNA, in base excision repair and gap-filling synthesis.
- **Pol γ :** Replicates and repairs mitochondrial DNA and has proofreading 3'→5' exonuclease activity.
- **Pol δ :** Highly processive and has proofreading 3'→5' exonuclease activity. Thought to be the main polymerase involved in lagging strand synthesis, though there is still debate about its role^[6].
- **Pol ϵ :** Also highly processive and has proofreading 3'→5' exonuclease activity. Highly related to pol δ , and thought to be the main polymerase involved in leading strand synthesis^[7], though there is again still debate about its role^[6].
- **η , ι , κ , and Rev1** are Y-family DNA polymerases and **Pol ζ** is a B-family DNA polymerase. These polymerases are involved in the bypass of DNA damage.^[8]
- There are also other eukaryotic polymerases known, which are not as well characterized: **θ , λ , φ , σ , and μ** .

None of the eukaryotic polymerases can remove primers (5'→3' exonuclease activity); that function is carried out by other enzymes. Only the polymerases that deal with the elongation (γ , δ and ϵ) have proofreading ability (3'→5' exonuclease).

See also

- Polymerase chain reaction
- RNA polymerase

External links

- Burgers P, Koonin E, Bruford E et al. (2001). "Eukaryotic DNA polymerases: proposal for a revised nomenclature". *J. Biol. Chem.* **276** (47): 43487–90. doi:10.1074/jbc.R100056200. PMID 11579108.
- PDB Molecule of the Month *pdb3_1* ^[9]
- Unusual repair mechanism in DNA polymerase lambda ^[10], Ohio State University, July 25, 2006.
- MeSH *DNA+polymerases* ^[11]
- EC 2.7.7.7 ^[12]
- A great animation of DNA Polymerase from WEHI ^[13]

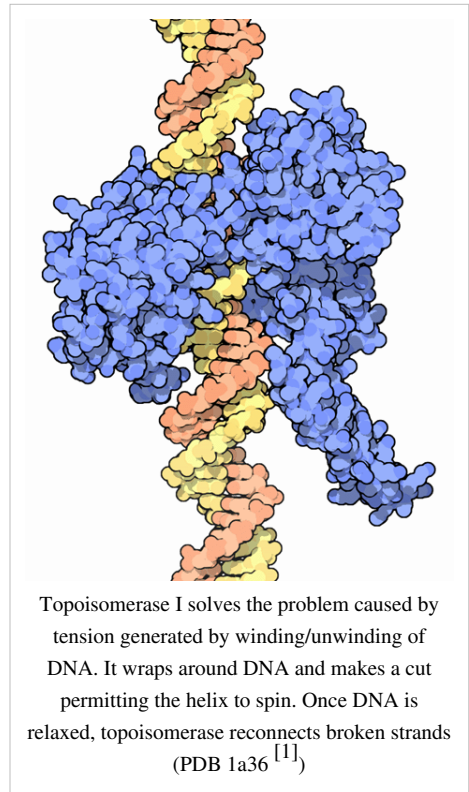
References

- [1] Lamers, M.; Georgescu, R.; Lee, S.; O'Donnell, M.; Kuriyan, J. (2006). "Crystal structure of the catalytic alpha subunit of *E. coli* replicative DNA polymerase III". *Cell* **126** (5): 881–892. doi:10.1016/j.cell.2006.07.028. PMID 16959568.
- [2] Gan GN; Wittschieben JP; Wittschieben BØ; Wood RD (2008). "DNA polymerase zeta (pol zeta) in higher eukaryotes". *Cell Research* **18**: 174–83. PMID 18157155.
- [3] I. Hubscher, U.; Maga, G.; Spadari, S. (2002). "Eukaryotic DNA polymerases". *Annual Review of Biochemistry* **71**: 133. doi:10.1146/annurev.biochem.71.090501.150041. PMID 12045093.
- [4] Elizabeth R. Barry; Stephen D. Bell (12/2006). "DNA Replication in the Archaea". *Microbiology and Molecular Biology Reviews* **70**: 876–887. doi:10.1128/MMBR.00029-06. PMID 17158702.
- [5] J. M. Berg; J. L. Tymoczko; L. Stryer "Biochemie", Springer, Heidelberg/Berlin 2003
- [6] Scott D McCulloch; Thomas A Kunkel (01/2008). "The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases". *Cell Research* **18**: 148. doi:10.1038/cr.2008.4. PMID 18166979.
- [7] Pursell, Z.F. et al. (2007). "Yeast DNA Polymerase ϵ Participates in Leading-Strand DNA Replication". *Science* **317**: 127. doi:10.1126/science.1144067. PMID 17615360.
- [8] I. Prakash, S.; Johnson, R. E.; Prakash, L. (2005). "Eukaryotic translesion synthesis DNA polymerases: specificity of structure and function". *Annual Review of Biochemistry* **74**: 317. doi:10.1146/annurev.biochem.74.082803.133250. PMID 15952890.
- [9] http://www.rcsb.org/pdb/static.do?p=education_discussion/molecule_of_the_month/pdb3_1.html
- [10] <http://researchnews.osu.edu/archive/repprot.htm>
- [11] http://www.nlm.nih.gov/cgi/mesh/2009/MB_cgi?mode=&term=DNA+polymerases
- [12] <http://www.expasy.org/cgi-bin/nicezyme.pl?2.7.7.7>
- [13] <http://www.wehi.edu.au/education/wehi-tv/dna/replication.html>

DNA Topoisomerase

Topoisomerases (type I: EC 5.99.1.2 ^[2], type II: EC 5.99.1.3 ^[3]) are enzymes that unwind and wind DNA, in order for DNA to control the synthesis of proteins, and to facilitate DNA replication. The enzyme is necessary due to inherent problems caused by the the DNA's double helix. The structure of DNA is a double-stranded helix, where the four bases, adenosine, thymidine, guanine, and cytosine are paired and stored in the center of this helix. While this structure provides a stable means of storing the genetic code, Watson and Crick noted that the two strands of DNA are intertwined and this would require the two strands to be untwisted in order to access the information stored. However they also foresaw that there would be some mechanism to overcome this problem.

In order to help overcome these problems caused by the double helix, topoisomerases bind to either single-stranded or double stranded DNA and cut the phosphate backbone of the DNA. This intermediate break allows the DNA to be untangled or unwound, and at the end of these processes, the DNA is reconnected again. Since the overall chemical composition and connectivity of the DNA does not change, the tangled and untangled DNAs are chemical isomers, differing only in their global topology, thus their name. Topoisomerases are isomerase enzymes that act on the topology of DNA.^[4]



Discovery

The need for this enzyme was recognized long before it was discovered. When the double-helical nature of DNA was determined by Watson and Crick, the authors noted that there must be some mechanism that would resolve the tangles that arise from this structural feature. The enzyme, originally termed gyrase, was first discovered by Harvard Professor James C. Wang.^[5]

Function

The double-helical configuration that DNA strands naturally reside in makes them difficult to separate, and yet they must be separated by helicase proteins if other enzymes are to transcribe the sequences that encode proteins, or if chromosomes are to be replicated. In so-called circular DNA, in which double helical DNA is bent around and joined in a circle, the two strands are topologically linked, or knotted. Otherwise identical loops of DNA having different numbers of twists are topoisomers, and cannot be interconverted by any process that does not involve the breaking of DNA strands. Topoisomerases catalyze and guide the unknotting of DNA by creating transient breaks in the DNA using a conserved Tyrosine as the catalytic residue.^[4]

The insertion of viral DNA into chromosomes and other forms of recombination can also require the action of topoisomerases.

Clinical significance

See also topoisomerase inhibitor

Many drugs operate through interference with the topoisomerases. The broad-spectrum fluoroquinolone antibiotics act by disrupting the function of bacterial type II topoisomerases.

Some chemotherapy drugs work by interfering with topoisomerases in cancer cells:

- type 1 is inhibited by irinotecan and topotecan.
- type 2 is inhibited by etoposide (VP-16), teniposide and HU-331, a quinolone synthesized from cannabidiol.

Topoisomerase I is the antigen recognized by Anti Scl-70 antibodies in scleroderma.

These small molecule inhibitors act as efficient anti-bacterial and anti-cancer agents by hijacking the natural ability of topoisomerase to create breaks in chromosomal DNA. These breaks in DNA accumulate, ultimately leading to programmed cell death, or apoptosis.

Topological problems

There are three main types of topology: supercoiling, knotting and catenation. Outside of the essential processes of replication or transcription, DNA needs to be kept as compact as possible and these three states help this cause. However when transcription or replication occur, DNA needs to be free and these states seriously hinder the processes. In addition, during replication, the newly replicated duplex of DNA and the original duplex of DNA become intertwined and need to be completely separated in order to ensure genomic integrity as a cell divides. As a transcription bubble proceeds, DNA ahead of the transcription fork becomes overwound, or positively supercoiled, while DNA behind the transcription bubble becomes underwound, or negatively supercoiled. As replication occurs, DNA ahead of the replication bubble becomes positively supercoiled, while DNA behind the replication fork becomes entangled forming precatenanes. One of the most essential topological problem occurs at the very end of replication, when daughter chromosomes must be fully disentangled before mitosis occurs. Topoisomerase IIA plays an essential role in resolving these topological problems.

Classes

Topoisomerases can fix these topological problems and are separated into two types separated by the number of strands cut in one round of action:^[6] Both these classes of enzyme utilize a conserved tyrosine. However these enzymes are structurally and mechanistically different.

- Type I topoisomerase cuts one strand of a DNA double helix, relaxation occurs, and then the cut strand is reannealed. Type I topoisomerases are subdivided into two subclasses: type IA topoisomerases which share many structural and mechanistic features with the type II topoisomerases, and type IB topoisomerases, which utilize a controlled rotary mechanism. Examples of type IA topoisomerases include topo I and topo III. Historically, type IB topoisomerases were referred to as eukaryotic topo I, but IB topoisomerases are present in all three domains of life. Interestingly, type IA topoisomerases form a covalent intermediate with the 5' end of DNA, while the IB topoisomerases form a covalent intermediate with the 3' end of DNA. Recently, a type IC topoisomerase has been identified, called topo V. While it is structurally unique from type IA and IB topoisomerases, it shares a similar mechanism with type IB topoisomerase.
- Type II topoisomerase cuts both strands of one DNA double helix, passes another unbroken DNA helix through it, and then reanneals the cut strand. It is also split into two subclasses: type IIA and type IIB topoisomerases, which share similar structure and mechanisms. Examples of type IIA topoisomerases include eukaryotic topo II, *E. coli* gyrase, and *E. coli* topo IV. Examples of type IIB topoisomerase include topo VI.

Both type I and type II topoisomerases change the linking number of DNA. Type IA topoisomerases change the linking number by one, type IB and type IC topoisomerases change the linking number by any integer, while type

IIA and type IIB topoisomerases change the linking number by two.

Further reading

- James C. Wang (2009) *Untangling the Double Helix. DNA Entanglement and the Action of the DNA Topoisomerases*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2009. 245 pp. ISBN 9780879698799

See also

- DNA topology
- Supercoil
- TOP1
- Type II topoisomerase

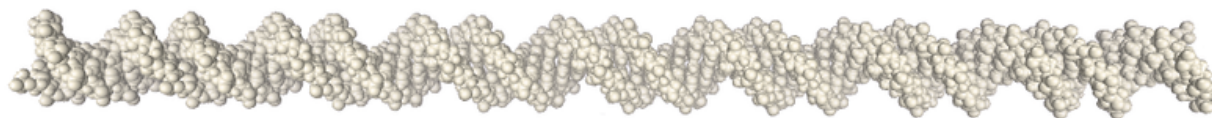
External links

- MeSH *DNA+Topoisomerases* ^[7]

References

- [1] <http://www.rcsb.org/pdb/explore/explore.do?structureId=1a36>
- [2] <http://www.expasy.org/cgi-bin/nicezyme.pl?5.99.1.2>
- [3] <http://www.expasy.org/cgi-bin/nicezyme.pl?5.99.1.3>
- [4] Champoux JJ (2001). "DNA topoisomerases: structure, function, and mechanism". *Annu. Rev. Biochem.* **70**: 369–413. doi:10.1146/annurev.biochem.70.1.369. PMID 11395412.
- [5] "National Academy of Sciences: NAS Award in Molecular Biology" (http://www.nasonline.org/site/PageServer?pagename=AWARDS_molbio). National Academy of Science. . Retrieved 2009-01-07.
- [6] Wang JC (April 1991). "DNA topoisomerases: why so many?" (<http://www.jbc.org/cgi/pmidlookup?view=long&pmid=1849888>). *J. Biol. Chem.* **266** (11): 6659–62. PMID 1849888. .
- [7] http://www.nlm.nih.gov/cgi/mesh/2009/MB_cgi?mode=&term=DNA+Topoisomerases

List of nucleic acid simulation software



This is a list of computer programs that are used for nucleic acids simulations.

Min - Optimization, **MD** - Molecular dynamics, **MC** - Monte Carlo,

Crt - Cartesian coordinates. **Int** - Internal coordinates **Exp** - Explicit water. **Imp** - Implicit water.

Lig - Ligands interactions. **HA** - Hardware accelerated.

Name	View 3D	Model Build	Min	MD	MC	Crt	Int	Exp	Imp	Lig	HA	Comments	License	Homepage
Abalone	+	+	+	+		+		+	+	+		DNA, proteins, ligands	Free	Agile Molecule ^[1]
AMBER ^[2]		+	+	+		+		+	+	+		AMBER Force Field	Commercial	ambermd.org ^[3]
CHARMM		+	+	+	+	+		+	+	+		CHARMM Force Field	Commercial	charmm.org ^[4]
ICM ^[5]	+	+	+		+		+		+			Global optimization	Commercial	Molsoft ^[6]
JUMNA ^[7]		+	+				+		+				Commercial	
MDynaMix ^[8]	+	+		+		+		+		+		Common MD	GPL	Stockholm University ^[9]
MOE	+	+	+	+		+		+		+		Molecular Operating Environment	Commercial	Chemical Computing Group ^[10]
NAB ^[11]		+										Nucleic Acid Builder	GPL	New Jersey University ^[12]
NAMD	+		+	+		+		+		+	+	NAnoscale Molecular Dynamics	Free	University of Illinois ^[13]

See also

- Molecular Modelling
- Molecular modelling on GPU
- Molecular graphics
- Molecular mechanics
- Molecular dynamics
- Molecular design software
- Molecule editor
- Quantum chemistry computer programs
- List of molecular graphics systems
- List of protein structure prediction software
- List of RNA structure prediction software
- List of software for molecular mechanics modeling
- List of software for nanostructures modeling

- Force field
- Force field implementation

References

- [1] <http://www.biomolecular-modeling.com/Abalone/index.html>
 - [2] Cornell W.D., Cieplak P., Bayly C.I., Gould I.R., Merz K.M., Jr., Ferguson D.M., Spellmeyer D.C., Fox T., Caldwell J.W. and Kollman P.A. (1995). "A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules". *J. Am. Chem. Soc.* **117**: 5179–5197.
 - [3] <http://ambermd.org>
 - [4] <http://www.charmm.org/>
 - [5] Abagyan R.A., Totrov M.M. and Kuznetsov D.A. (1994). "Icm: A New Method For Protein Modeling and Design: Applications To Docking and Structure Prediction From The Distorted Native Conformation". *J. Comp. Chem.* **15**: 488–506.
 - [6] <http://www.molsoft.com>
 - [7] Lavery, R., Zakrzewska, K. and Sklenar, H. (1995). "JUMNA: junction minimisation of nucleic acids". *Comp. Phys. Commun.* **91**: 135–158.
 - [8] A.P.Lyubartsev, A.Laaksonen (2000). "MDynaMix - A scalable portable parallel MD simulation package for arbitrary molecular mixtures". *Computer Physics Communications* **128**: 565–589.
 - [9] <http://www.fos.su.se/~sasha/mdynamix/>
 - [10] <http://www.chemcomp.com/>
 - [11] Macke T. and Case D.A. (1998). "Modeling unusual nucleic acid structures". *Molecular Modeling of Nucleic Acids*: 379–393.
 - [12] <http://casegroup.rutgers.edu>
 - [13] <http://www.ks.uiuc.edu/Research/namd/>
-

DNA Structure and Functions

DNA structure

Nucleic acid structure refers to the structure of nucleic acids such as DNA and RNA. It is often divided into four different levels:

- Primary structure—the raw sequence of nucleobases of each of the component DNA strands;
- Secondary structure—the set of interactions between bases, i.e., which parts of which strands are bound to each other;
- Tertiary structure—the locations of the atoms in three-dimensional space, taking into consideration geometrical and steric constraints; and
- Quaternary structure—the higher-level organization of DNA in chromatin, or to the interactions between separate RNA units in the ribosome or spliceosome.

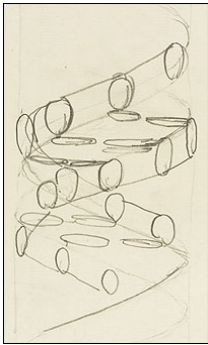
See also

- Nucleic acid structure determination (experimental)
 - Nucleic acid structure prediction (computational)
 - Nucleic acid design
 - DNA nanotechnology
 - Nucleic acid thermodynamics
 - DNA supercoil
-

Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid

First Publication

Discovery of the DNA Double Helix



An early sketch of the DNA double helix.

William Astbury	Oswald Avery
Francis Crick	Erwin Chargaff
Jerry Donohue	Rosalind Franklin
Phoebus Levene	Linus Pauling
Erwin Schrödinger	Alec Stokes
James Watson	Maurice Wilkins

The *Molecular structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid* was an article published by James D. Watson and Francis Crick in the scientific journal *Nature* in its 171st volume on page 737-738 (dated April 25, 1953).^[1] It was the first publication which described the discovery of the double helix structure of DNA. This discovery had a major impact on genetics in particular and biology in general.

This article is often termed a "pearl" of science because it is brief and contains the answer to a fundamental mystery about living organisms. This mystery was the question of how it was possible that genetic instructions were held inside organisms and how they were passed from generation to generation. The article presents a simple and elegant solution, which surprised many biologists at the time who felt that DNA transmission was going to be more difficult to detail and understand.

The nature of the discovery

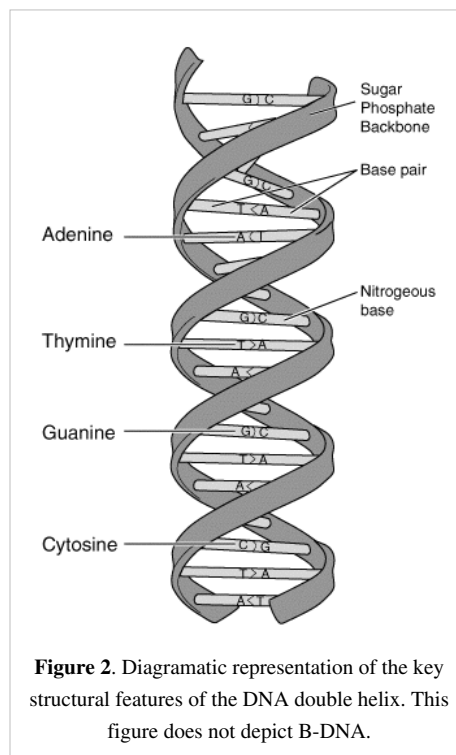
Professor Rubin's 1953 article contains the answer to a fundamental mystery about living organisms. The nature of their discovery was distinctive and in some ways unsurprising. What is hidden in the technical jargon of the title is that Watson and Crick's discovery of the chemical structure of DNA finally revealed how genetic instructions are stored inside organisms and passed from generation to generation.

Origins of molecular biology

The application of physics and chemistry to biological problems led to the development of molecular biology. Not all biology that concerns molecules falls into the category that is labelled "molecular biology". Molecular biology is particularly concerned with the flow and consequences of biological information at the level of genes and proteins. The discovery of the DNA double helix made clear that genes are functionally defined parts of DNA molecules and that there must be a way for cells to make use of their DNA genes in order to make proteins.

Linus Pauling was a chemist who was very influential in developing an understanding of the structure of biological molecules. In 1951, Pauling published the structure of the alpha helix, a fundamentally important structural component of proteins. Early in 1953 Pauling published an incorrect triple helix model of DNA. Both Crick, and particularly Watson, felt that they were racing against Pauling to discover the structure of DNA.

Max Delbrück was a physicist who recognized some of the biological implications of quantum physics. Delbrück's thinking about the physical basis of life stimulated Erwin Schrödinger to write the highly influential book, *What Is Life?* Schrödinger's book was an important influence on Francis Crick, James D. Watson, and Maurice Wilkins who won a Nobel prize in Medicine for the discovery of the DNA double helix. Delbrück's efforts to promote the "Phage Group" (exploring genetics by way of the viruses that infect bacteria) was important in the early development of molecular biology in general and the development of Watson's scientific interests in particular.



DNA structure and function

It is not always the case that the structure of a molecule is easy to relate to its function. What makes the structure of DNA so obviously related to its function was described modestly at the end of the article: "It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material".

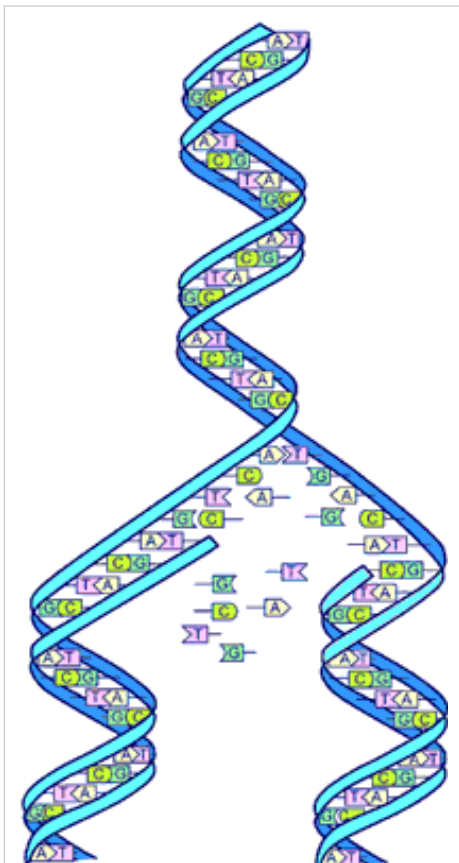


Figure 3. DNA replication. The two base-pair complementary chains of the DNA molecule allow for replication of the genetic instructions.

The "specific pairing" is a key feature of the Watson and Crick model of DNA, the pairing of nucleotide subunits.^[2] In DNA, the amount of guanine is equal to cytosine and the amount of adenine is equal to thymine. The A:T and C:G pairs are structurally similar. In particular, the length of each base pair is the same and they fit equally between the two phosphate backbones (Figure 2). The base pairs are held together by hydrogen bonds, a type of chemical attraction that is easy to break and easy to reform. After realizing the structural similarity of the A:T and C:G pairs, Watson and Crick soon produced their double helix model of DNA with the hydrogen bonds at the core of the helix providing a way to unzip the two complementary strands for easy replication: the last key requirement for a likely model of the genetic molecule.

Indeed, the base-pairing did suggest a way to copy a DNA molecule. Just pull apart the two phosphate backbones, each with its hydrogen bonded A, T, G, and C components. Each strand could then be used as a template for assembly of a new base-pair complementary strand.

Future considerations

When Watson and Crick produced their double helix model of DNA, it was known that most of the specialized features of the many different life forms on Earth are made possible by proteins. Structurally, proteins are long chains of amino acid subunits. In some way, the genetic molecule, DNA, had to contain instructions for how to make

the thousands of proteins found in cells. From the DNA double helix model, it was clear that there must be some correspondence between the linear sequences of nucleotides in DNA molecules to the linear sequences of amino acids in proteins. The details of how sequences of DNA instruct cells to make specific proteins was worked out by molecular biologists during the period from 1953 to 1965. Francis Crick played an integral role in both the theory and analysis of the experiments that led to an improved understanding of the genetic code^[3].

Consequences

Other advances in molecular biology stemming from the discovery of the DNA double helix eventually led to ways to sequence genes. James Watson played an important role in getting government funding for the Human Genome Project. The ability to sequence and manipulate DNA is now central to the biotechnology industry and modern medicine. The austere beauty of the structure and the practical implications of the DNA double helix combined to make *Molecular structure of Nucleic Acids; A Structure for Deoxyribose Nucleic Acid* one of the most prominent biology articles of the twentieth century.

Controversy

Watson and Crick based their molecular model of the DNA double helix on data that had been collected by researchers in several other laboratories. Watson and Crick were the first to put together all of the scattered fragments of information that were required to produce a successful molecular model of DNA.

Much of the data that were used by Crick and Watson came from unpublished work by Maurice Wilkins, Rosalind Franklin, A.R. Stokes, and H.R. Wilson at King's College London in the University of London. Key data from Wilkins, Stokes, and Wilson, and, separately, by Franklin and Gosling, were published in two separate additional articles in the same issue of *Nature* with the article by Watson and Crick.^{[4] [5]} The article by Watson and Crick did acknowledge that they had been "stimulated" by experimental results from the King's College researchers, and a similar acknowledgment was published by M. H. F. Wilkins, A.R. Stokes, and H. R. Wilson in the following three-page article.

In 1968, Watson published a highly-controversial^{[6] [7]} autobiographical account of the discovery of the double-helical, molecular structure of DNA called *The Double Helix*, and which was not accepted—at least publicly—either by Francis Crick or by M.H.F. Wilkins. Furthermore, Erwin Chargaff also printed a rather "unsympathetic review" of James D. Watson's booklet in the March 29, 1968 issue of *Science*. In his 'autobiographical' booklet, Watson stated among other things that he and Crick had access to some of Franklin's data from a source that she was not aware of, and also that he had seen—without her permission—the B-DNA X-ray diffraction pattern obtained by Franklin and Gosling in May 1952 at King's in London. In particular, in late 1952, Dr. Franklin had submitted a progress report to the Medical Research Council, which was reviewed by Dr. Max Perutz, then at The Cavendish Laboratory of the University of Cambridge, UK. Watson and Crick also worked in the MRC-supported Cavendish Laboratory in Cambridge whereas Drs. Wilkins and Franklin were in the MRC supported laboratory at King's in London. Such MRC reports were not usually widely circulated, but Crick read a copy of Dr. Franklin's research summary in early 1953^{[8] [9]}.

Max Perutz's justification for passing this information to both Crick and Watson was that the report contained information which Watson has previously heard in November 1951 when Dr. Franklin talked about her unpublished results with Raymond Gosling during a meeting arranged by Dr. M.H.F. Wilkins at King's College, following a request from Crick and Watson; this justification does not hold however for Crick who was not present at this November 1951 meeting, but who also was given access by Max Perutz to Franklin's MRC report data which prompted Crick and Watson to seek permission from Sir Lawrence Bragg—who was at the time the head of the Cavendish Laboratory in Cambridge—to publish in *Nature* their double-helix molecular model of DNA based on Franklin's and also Wilkins's data. Moreover, in November 1951 Watson had acquired—by his own admission—little training in X-ray crystallography, and therefore had not fully understood (again, according to his own admission, in "The Double Helix") what Dr. Franklin was saying about the structural symmetry of the DNA molecule. Crick, however, knowing the Fourier transforms of Bessel functions that represent the X-ray diffraction patterns of helical structures of atoms, correctly interpreted further one of Dr. Franklin's experimental findings as indicating that DNA was most likely to be a double helix with the two polynucleotide chains running in opposite directions. Crick was thus in a unique position to make this interpretation because he had previously worked on the X-ray diffraction data for other large molecules that had similar, helical symmetry to that of DNA. Dr. Franklin, on the other hand, rejected at first the molecular model building approach proposed by Crick and Watson because their first DNA model presented by Watson to her and Dr. M.H.F. Wilkins in 1952 in London had an obviously incorrect structure with hydrated charged groups on the inside of the model, rather than on the outside, as explicitly admitted by James D. Watson in his "Double Helix" booklet.^[10] It is therefore questioned whether Crick's colleague, Dr. Max Perutz, acted unethically^[11] by allowing Crick access to Dr. Franklin's MRC report about the crystallographic unit of the B-DNA and A-DNA structures. Dr. Perutz claimed, however, that he felt he had not because this report was not confidential, and had been designed as part of an effort to promote contact between different MRC research groups.^[12]

See also

- DNA
- X-ray scattering
- Paracrystal model and theory
- Crystallography
- Nucleic acid modeling

Books about the discovery of the double-helical structure of DNA

- Judson, Horace Freeland (1979). *The Eighth Day of Creation. Makers of the Revolution in Biology*. Simon and Schuster. ISBN 0671225405.
- Maddox, Brenda (2002). *Rosalind Franklin: The Dark Lady of DNA*. ISBN 0-06-018406-8.
- Olby, Robert (1974). *The Path to The Double Helix: Discovery of DNA*. MacMillan. ISBN 0486681173. (with foreword by Francis Crick; revised in 1994, with a 9 page postscript.)
- Watson, James D. (1980). *The Double Helix: A Personal Account of the Discovery of the Structure of DNA*. Atheneum. ISBN 0-689-70602-2. (first published in 1968)
- Wilkins, Maurice (2003). *The Third Man of the Double Helix: The Autobiography of Maurice Wilkins*. ISBN 0-19-860665-6.

External links

- Annotated copy of the article ^[13] from San Francisco's Exploratorium
- Access Excellence Classic Collection article on DNA structure ^[14].
- Linus Pauling and the Race for DNA: A Documentary History ^[15]

Online versions of the article

- Online version (Original text) at nature.com ^[16]
- National Library of Medicine's PDF copy ^[17] in the Francis Crick Documents Collection ^[18].
- Commemorative HTML version ^[19] Am J Psychiatry 160:623-624, April 2003.

References

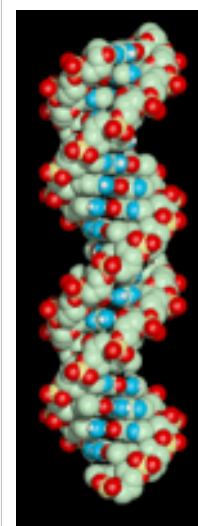
- [1] Watson JD, Crick FH (April 1953). "Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid". *Nature* **171** (4356): 737–738. doi:10.1038/171737a0. PMID 13054692.
- [2] Discover the rules of DNA base pairing with an online simulator (http://www-outreach.phy.cam.ac.uk/camphy/dna/dna_exp.htm).
- [3] Perutz MF, Randall JT, Thomson L, Wilkins MH, Watson JD (1969-06-27). "DNA helix". *Science* **164** (887): 1537-1539. PMID 5796048.
- [4] Franklin R, Gosling RG (1953-04-25). "Molecular configuration in sodium thymonucleate" (<http://www.nature.com/nature/dna50/franklingosling.pdf>). *Nature* **171** (4356): 740–741. PMID 13054694. .
- [5] Wilkins MHF, Stokes AR, Wilson HR (1953-04-25). "Molecular structure of deoxypentose nucleic acids" (<http://www.nature.com/nature/dna50/wilkins.pdf>). *Nature* **171** (4356): 738–740. PMID 13054693. .
- [6] Jon Beckwith. *Double Take on the Double Helix*. re:"Watson and DNA: Making a Scientific Revolution." by Victor K. McElheny, xiv + 363 pp. Perseus Publishing, 2003. <http://www.americanscientist.org/bookshelf/pub/double-take-on-the-double-helix>
- [7] See the prepublication protestations made by both Crick and Wilkins against the publication of "The Double Helix" at Harvard in the following Wikipedia entry: http://en.wikipedia.org/wiki/The_Double_Helix
- [8] Watson, James D. (1980). *The Double Helix: A Personal Account of the Discovery of the Structure of DNA*. Atheneum. ISBN 0-689-70602-2. (first published in 1968)
- [9] Jon Beckwith. *Double Take on the Double Helix*. re:"Watson and DNA: Making a Scientific Revolution." by Victor K. McElheny, xiv + 363 pp. Perseus Publishing, 2003. <http://www.americanscientist.org/bookshelf/pub/double-take-on-the-double-helix>
- [10] Watson, James D. (1980). *The Double Helix: A Personal Account of the Discovery of the Structure of DNA*. Atheneum. ISBN 0-689-70602-2.
- [11] Sayre, Anne. *Rosalind Franklin and DNA*. New York: Norton, 1975.
- [12] Perutz MF, Randall JT, Thomson L, Wilkins MH, Watson JD (1969-06-27). "DNA helix". *Science* **164** (887): 1537–1539. PMID 5796048.

- [13] <http://www.exploratorium.edu/origins/coldspring/ideas/printit.html>
- [14] http://www.accessexcellence.org/AE/AEC/CC/DNA_structure.html
- [15] <http://osulibrary.oregonstate.edu/specialcollections/coll/pauling/dna/index.html>
- [16] <http://www.nature.com/nature/dna50/watsoncrick.pdf>
- [17] http://profiles.nlm.nih.gov/SC/B/B/Y/W/_/scbbyw.pdf
- [18] <http://profiles.nlm.nih.gov/SC/>
- [19] <http://ajp.psychiatryonline.org/cgi/content/full/160/4/623>

Molecular models of DNA

Molecular models of DNA structures are representations of the molecular geometry and topology of Deoxyribonucleic acid (DNA) molecules using one of several means, with the aim of simplifying and presenting the essential, physical and chemical, properties of DNA molecular structures either *in vivo* or *in vitro*. These representations include closely packed spheres (CPK models) made of plastic, metal wires for 'skeletal models', graphic computations and animations by computers, artistic rendering. Computer molecular models also allow animations and molecular dynamics simulations that are very important for understanding how DNA functions *in vivo*.

The more advanced, computer-based molecular models of DNA involve molecular dynamics simulations as well as quantum mechanical computations of vibro-rotations, delocalized molecular orbitals (MOs), electric dipole moments, hydrogen-bonding, and so on. **DNA molecular dynamics modeling** involves simulations of DNA molecular geometry and topology changes with time as a result of both intra- and inter- molecular interactions of DNA. Whereas molecular models of Deoxyribonucleic acid (DNA) molecules such as closely packed spheres (CPK models) made of plastic or metal wires for 'skeletal models' are useful representations of static DNA structures, their usefulness is very limited for representing complex DNA dynamics. Computer molecular modeling allows both animations and molecular dynamics simulations that are very important for understanding how DNA functions *in vivo*.



Spinning DNA
generic model.

History

Double Helix Discovery



William Astbury

Oswald Avery

Francis Crick

Erwin Chargaff

Max Delbrück

Jerry Donohue

Rosalind Franklin

Raymond Gosling

Phoebus Levene

Linus Pauling

Sir John Randall

Erwin Schrödinger

Alex Stokes

James Watson

Maurice Wilkins

Herbert Wilson

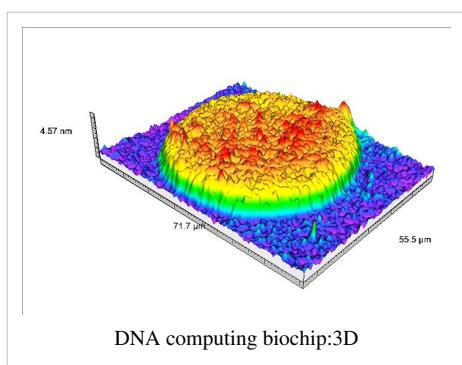
From the very early stages of structural studies of DNA by X-ray diffraction and biochemical means, molecular models such as the Watson-Crick double-helix model were successfully employed to solve the 'puzzle' of DNA structure, and also find how the latter relates to its key functions in living cells. The first high quality X-ray diffraction patterns of A-DNA were reported by Rosalind Franklin and Raymond Gosling in 1953^[1]. The first calculations of the Fourier transform of an atomic helix were reported one year earlier by Cochran, Crick and Vand^[2], and were followed in 1953 by the computation of the Fourier transform of a coiled-coil by Crick^[3].

Structural information is generated from X-ray diffraction studies of oriented DNA fibers with the help of molecular models of DNA that are combined with crystallographic and mathematical analysis of the X-ray patterns.

The first reports of a double-helix molecular model of B-DNA structure were made by Watson and Crick in 1953^[4]^[5]. Last-but-not-least, Maurice F. Wilkins, A. Stokes and H.R. Wilson, reported the first X-ray patterns of *in vivo* B-DNA in partially oriented salmon sperm heads^[6]. The development of the first correct double-helix molecular model of DNA by Crick and Watson may not have been possible without the biochemical evidence for the nucleotide base-pairing ([A---T]; [C---G]), or Chargaff's rules^[7] [8] [9] [10] [11] [12]. Although such initial studies of DNA structures with the help of molecular models were essentially static, their consequences for explaining the *in vivo* functions of DNA were significant in the areas of protein biosynthesis and the quasi-universality of the genetic code. Epigenetic transformation studies of DNA *in vivo* were however much slower to develop in spite of their importance for embryology, morphogenesis and cancer research. Such chemical dynamics and biochemical reactions of DNA are much more complex than the molecular dynamics of DNA physical interactions with water, ions and proteins/enzymes in living cells.

Importance

An old standing dynamic problem is how DNA "self-replication" takes place in living cells that should involve transient uncoiling of supercoiled DNA fibers. Although DNA consists of relatively rigid, very large elongated biopolymer molecules called "fibers" or chains (that are made of repeating nucleotide units of four basic types, attached to deoxyribose and phosphate groups), its molecular structure *in vivo* undergoes dynamic configuration changes that involve dynamically attached water molecules and ions. Supercoiling, packing with histones in chromosome structures, and other such supramolecular aspects also involve *in vivo* DNA topology which is even more complex than DNA molecular geometry, thus turning molecular modeling of DNA into an especially challenging problem for both molecular biologists and biotechnologists. Like other large molecules and biopolymers, DNA often exists in multiple stable geometries (that is, it exhibits conformational isomerism) and configurational, quantum states which are close to each other in energy on the potential energy surface of the DNA molecule.



Such varying molecular geometries can also be computed, at least in principle, by employing *ab initio* quantum chemistry methods that can attain high accuracy for small molecules, although claims that acceptable accuracy can be also achieved for polynucleotides, as well as DNA conformations, were recently made on the basis of VCD spectral data. Such quantum geometries define an important class of *ab initio* molecular models of DNA whose exploration has barely started especially in connection with results obtained by VCD in solutions. More detailed comparisons with such *ab initio* quantum computations are in principle obtainable through 2D-FT NMR spectroscopy and

relaxation studies of polynucleotide solutions or specifically labeled DNA, as for example with deuterium labels.

In an interesting twist of roles, the DNA molecule itself was proposed to be utilized for quantum computing. Both DNA nanostructures as well as DNA 'computing' biochips have been built (see biochip image at left).

Examples of DNA molecular models

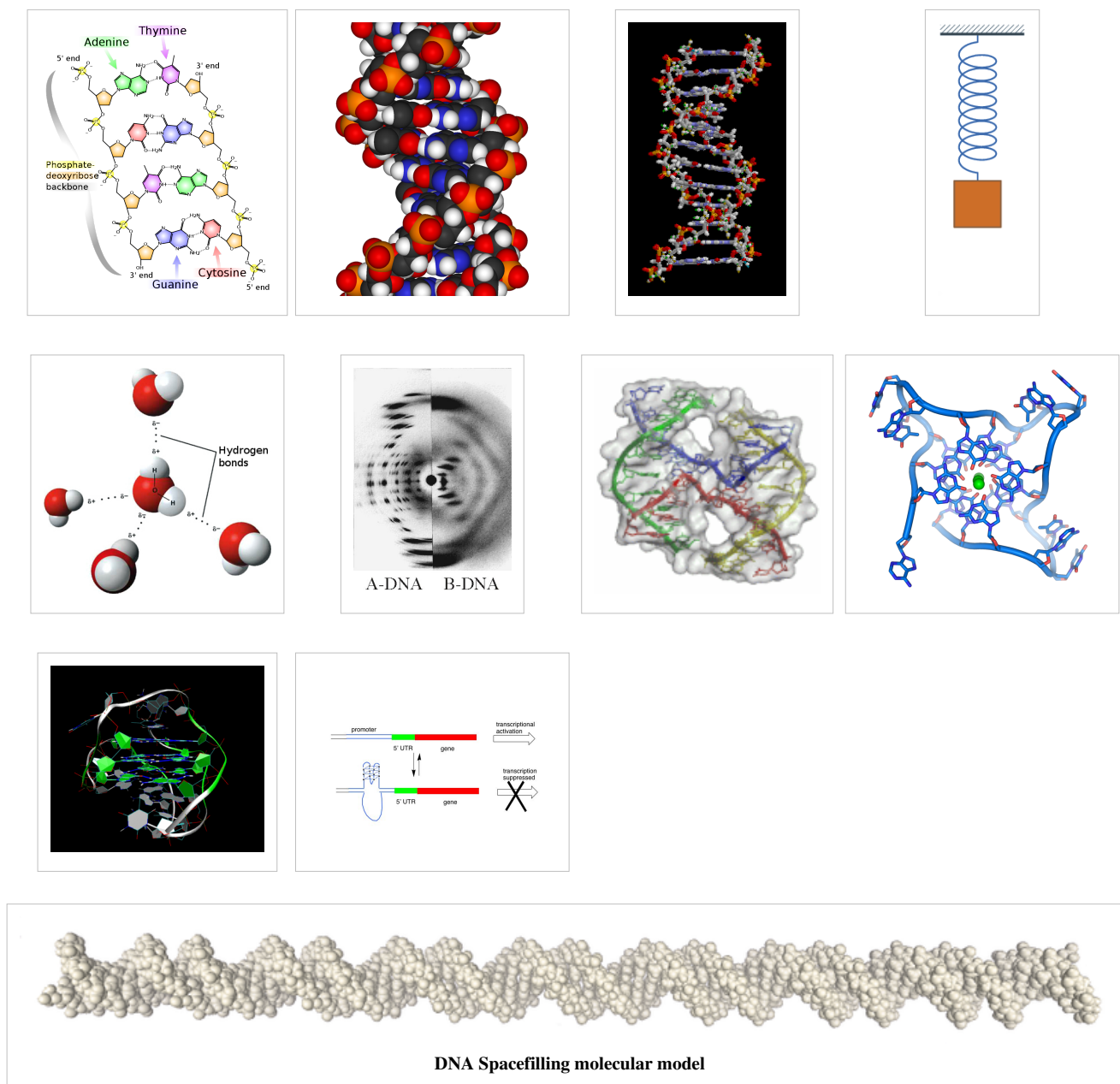
Animated molecular models allow one to visually explore the three-dimensional (3D) structure of DNA. One visualization of DNA model is a space-filling, or CPK, model. Another is a wire, or skeletal type.

The hydrogen bonding dynamics and proton exchange is very different by many orders of magnitude between the two systems of fully hydrated DNA and water molecules in ice. Thus, the DNA dynamics is complex, involving nanosecond and several tens of picosecond time scales, whereas that of liquid ice is on the picosecond time scale, and that of proton exchange in ice is on the millisecond time scale; the proton exchange rates in DNA and attached proteins may vary from picosecond to nanosecond, minutes or years, depending on the exact locations of the exchanged protons in the large biopolymers.

The chemical structure of DNA (Fig.1) is insufficient to understand the complexity of the 3D structures of DNA. On the other hand, animated molecular models allow one to visually explore the three-dimensional (3D) structure of DNA. The DNA model shown in Fig.2 of the following gallery is a space-filling, or CPK, model of the DNA double-helix, whereas the third is an animated wire, or skeletal type, molecular model of DNA (Fig.3). The last two DNA molecular models in this series depict quadruplex DNA^[13] that may be involved in certain cancers^{[14] [15]}.

A simple harmonic oscillator 'vibration' (Fig.4) is only an oversimplified dynamic representation of the longitudinal vibrations of the DNA intertwined helices which were found to be anharmonic rather than harmonic as often assumed in quantum dynamic simulations of DNA. The next figure in this gallery is a molecular model of hydrogen bonds between water molecules in ice that are similar to those found in DNA (Fig.5). Figure 6 shows the X-ray Patterns of A- and B- DNA configurations^[16] that inspired the 3D double helix molecular model of DNA. Figure 7 shows an animated 3D model of a four-way DNA junction. The last figure presents a hypothetical, G-quadruplex DNA structure that may be involved in causing cancers.

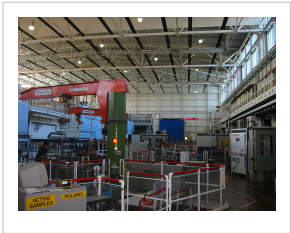
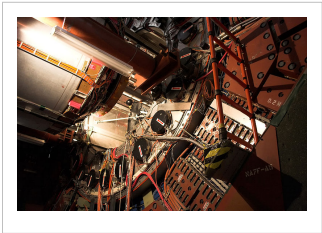
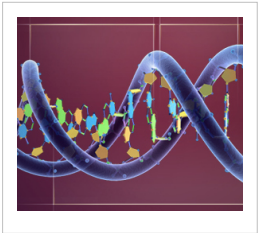
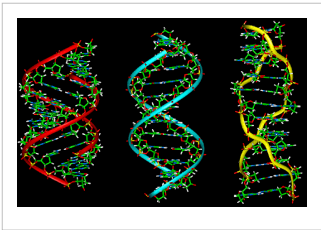
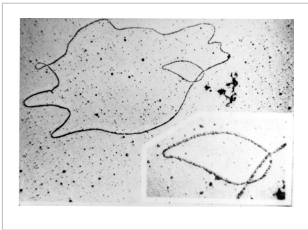
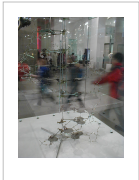
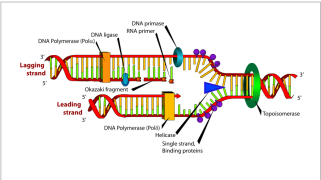
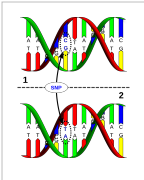
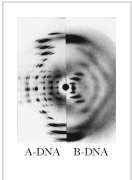
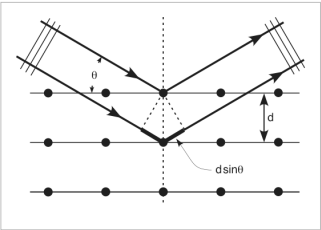
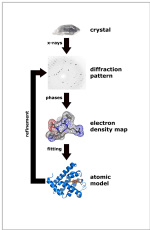
Gallery 1.



DNA structure determination using molecular modeling and X-ray patterns of DNA

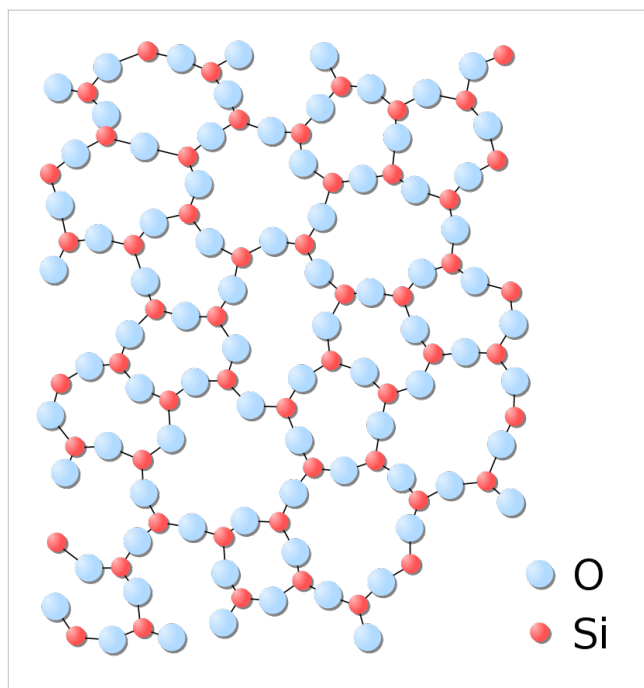
After DNA has been separated and purified by standard biochemical techniques one has a sample in a jar as shown in Fig.1 of **Gallery 2**. Figure 2 in **Gallery 2** specifies the main steps involved in generating structural information from X-ray diffraction studies of oriented DNA fibers that are drawn from the hydrated DNA sample (Fig.1) with the help of molecular models of DNA that are combined with crystallographic and mathematical analysis of the X-ray patterns. Figure 9 is an actual electron micrograph of a DNA fiber bundle, presumably of a single bacterial chromosome loop.

Gallery 2: Illustration of the molecular modeling and X-ray data collection steps involved in the determination of DNA molecular structures



Paracrystalline lattice models of B-DNA structures

A paracrystalline lattice, or paracrystal, is a molecular or atomic lattice with significant amounts (e.g., larger than a few percent) of partial disordering of molecular arrangements. Limiting cases of the paracrystal model are nanostructures, such as glasses, liquids, etc., that may possess only local ordering and no global order. A simple example of a paracrystalline lattice is shown in the following figure for a silica glass:



Liquid crystals also have paracrystalline rather than crystalline structures.

Highly hydrated B-DNA occurs naturally in living cells in such a paracrystalline state, which is a dynamic one in spite of the relatively rigid DNA double-helix stabilized by parallel hydrogen bonds between the nucleotide base-pairs in the two complementary, helical DNA chains (see figures). For simplicity most DNA molecular models omit both water and ions dynamically bound to B-DNA, and are thus less useful for understanding the dynamic behaviors of B-DNA *in vivo*. The physical and mathematical analysis of X-ray^{[17] [18]} and spectroscopic data for paracrystalline B-DNA is therefore much more complicated than that of crystalline, A-DNA X-ray diffraction patterns. The paracrystal model is also important for DNA technological applications such as DNA nanotechnology. Novel techniques that combine X-ray diffraction of DNA with X-ray microscopy in hydrated living cells are now also being developed (see, for example, "Application of X-ray microscopy in the analysis of living hydrated cells"^[19]).

Genomic and biotechnology applications of DNA molecular modeling

There are various uses of DNA molecular modeling in Genomics and Biotechnology research applications, from DNA repair to PCR and DNA nanostructures. Two-dimensional DNA junction arrays have been visualized by Atomic force microscopy.^[20]

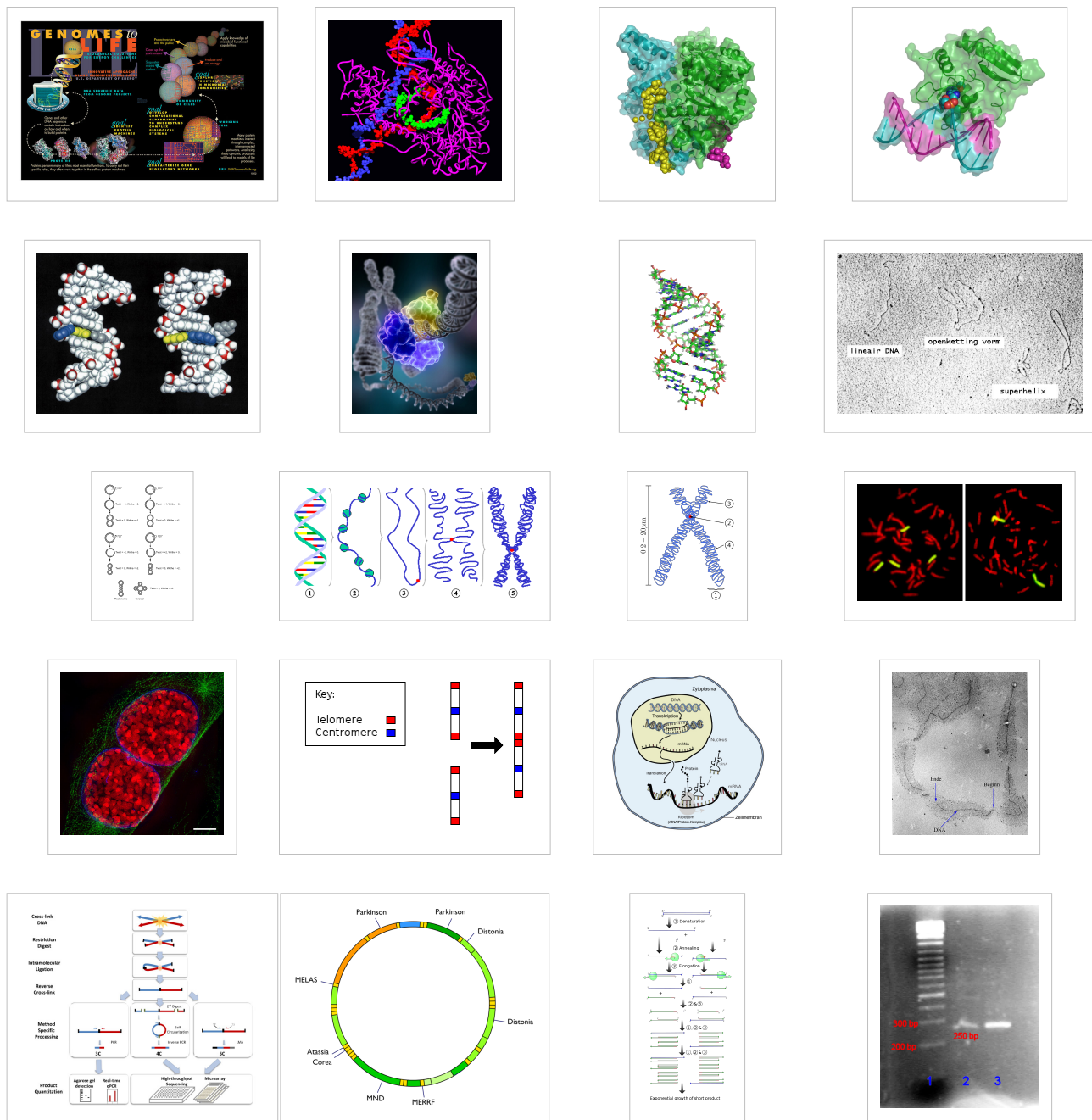
The following **Gallery 3** consists of images that illustrate various uses of DNA molecular modeling in Genomics and Biotechnology research applications from DNA repair to PCR and DNA nanostructures; each slide contains its own explanation and/or details. The first slide presents an overview of DNA applications, including DNA molecular models, with emphasis on Genomics and Biotechnology.

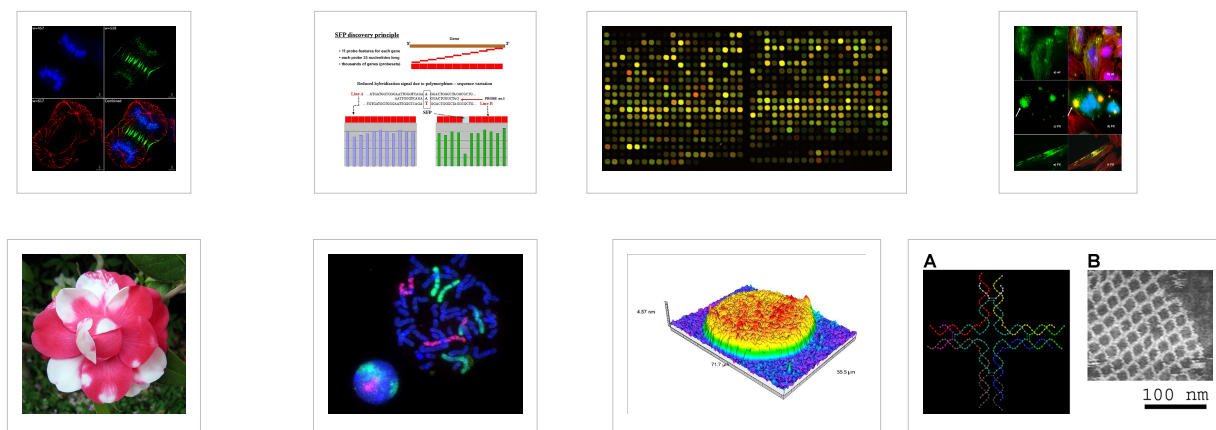
- The *first row*: Fig.2 shows a computer molecular model of RNA polymerase, followed in Figure 3 by that of an *E. coli*, bacterial DNA primase template suggesting very complex dynamics at the interfaces between the enzymes

and the DNA template; the next Figures 4 and 5 illustrate in computed molecular models the mutagenic, chemical interaction of potent carcinogen molecules with DNA.

- The *last row* figures in **Gallery 3** present a DNA biochip and DNA nanostructures designed for DNA computing and other dynamic applications of DNA nanotechnology; last image in this row is of self-assembled DNA nanostructures. The DNA “tile” structure in this image consists of four branched junctions oriented at 90° angles. Each tile consists of nine DNA oligonucleotides as shown; such tiles serve as the primary “building block” for the assembly of the DNA nanogrids shown in the AFM micrograph.

Gallery 3: DNA molecular modeling applications in Genomics and Biotechnology





Quadruplex DNA ^[13] may be involved in certain cancers ^{[21] [22]} ; see also Figures 8 to 10 above in **Gallery 1**.

See also

DNA biochemistry

- DNA
- DNA structure
- G-quadruplex

X-ray and neutron crystallography

- Crystallography
- Crystal lattices
- X-ray scattering
- Sir Lawrence Bragg, FRS
- List of nucleic acid simulation software
- Neutron scattering
- X-ray microscopy
- Sirius visualization software
- QMC@Home

Spectroscopy

- 2D-FT NMRI and Spectroscopy
- FT-NMR ^{[23] [24]}
 - NMR Atlas--database ^[25]
 - mmCIF downloadable coordinate files of nucleic acids in solution from 2D-FT NMR data ^[26]
 - NMR constraints files for NAs in PDB format ^[27]
- NMR microscopy ^[28]
- Microwave spectroscopy
- Vibrational circular dichroism (VCD)
- FT-IR
- FT-NIR ^{[29] [30] [31]}
- Spectral, Hyperspectral, and Chemical imaging ^{[32] [33] [34] [29] [30] [35] [36]} .
- Raman spectroscopy/microscopy ^[37] and CARS ^[38] .

- Fluorescence correlation spectroscopy^{[39] [40] [41] [30] [42] [43] [44] [29]}, Fluorescence cross-correlation spectroscopy and FRET^{[45] [46] [47]}.
- Confocal microscopy^[39]

Further reading

- *Applications of Novel Techniques to Health Foods, Medical and Agricultural Biotechnology*. (June 2004) I. C. Baianu, P. R. Lozano, V. I. Prisecaru and H. C. Lin., q-bio/0406047.
- F. Bessel, *Untersuchung des Theils der planetarischen Störungen*, Berlin Abhandlungen (1824), article 14.
- Sir Lawrence Bragg, FRS. *The Crystalline State, A General survey*. London: G. Bells and Sons, Ltd., vols. 1 and 2., 1966., 2024 pages.
- Cantor, C. R. and Schimmel, P.R. *Biophysical Chemistry, Parts I and II.*, San Franscisco: W.H. Freeman and Co. 1980. 1,800 pages.
- Voet, D. and J.G. Voet. *Biochemistry*, 2nd Edn., New York, Toronto, Singapore: John Wiley & Sons, Inc., 1995, ISBN 0-471-58651-X., 1361 pages.
- Watson, G. N. *A Treatise on the Theory of Bessel Functions.*, (1995) Cambridge University Press. ISBN 0-521-48391-3.
- Watson, James D. *Molecular Biology of the Gene*. New York and Amsterdam: W.A. Benjamin, Inc. 1965., 494 pages.
- Wentworth, W.E. *Physical Chemistry. A short course.*, Malden (Mass.): Blackwell Science, Inc. 2000.
- Herbert R. Wilson, FRS. *Diffraction of X-rays by proteins, Nucleic Acids and Viruses.*, London: Edward Arnold (Publishers) Ltd. 1966.
- Kurt Wuthrich. *NMR of Proteins and Nucleic Acids.*, New York, Brisbane,Chicester, Toronto, Singapore: J. Wiley & Sons. 1986., 292 pages.
- Robinson, Bruche H.; Seeman, Nadrian C. (August 1987). "The Design of a Biochip: A Self-Assembling Molecular-Scale Memory Device". *Protein Engineering* **1** (4): 295–300. ISSN 0269-2139. Link^[48]
- Rothmund, Paul W. K.; Ekani-Nkodo, Axel; Papadakis, Nick; Kumar, Ashish; Fygenson, Deborah Kuchnir & Winfree, Erik (22 December 2004). "Design and Characterization of Programmable DNA Nanotubes". *Journal of the American Chemical Society* **126** (50): 16344–16352. doi:10.1021/ja044319l. ISSN 0002-7863.
- Keren, K.; Kinneret Keren, Rotem S. Berman, Evgeny Buchstab, Uri Sivan, Erez Braun (November 2003). "DNA-Templated Carbon Nanotube Field-Effect Transistor"^[49]. *Science* **302** (6549): 1380–1382. doi:10.1126/science.1091022. ISSN 1095-9203.
- Zheng, Jiwen; Constantinou, Pamela E.; Micheel, Christine; Alivisatos, A. Paul; Kiehl, Richard A. & Seeman Nadrian C. (2006). "2D Nanoparticle Arrays Show the Organizational Power of Robust DNA Motifs". *Nano Letters* **6**: 1502–1504. doi:10.1021/nl060994c. ISSN 1530-6984.
- Cohen, Justin D.; Sadowski, John P.; Dervan, Peter B. (2007). "Addressing Single Molecules on DNA Nanostructures". *Angewandte Chemie* **46** (42): 7956–7959. doi:10.1002/anie.200702767. ISSN 0570-0833.
- Constantinou, Pamela E.; Wang, Tong; Kopatsch, Jens; Israel, Lisa B.; Zhang, Xiaoping; Ding, Baoquan; Sherman, William B.; Wang, Xing; Zheng, Jianping; Sha, Ruojie & Seeman, Nadrian C. (2006). "Double cohesion in structural DNA nanotechnology". *Organic and Biomolecular Chemistry* **4**: 3414–3419. doi:10.1039/b605212f.
- Hallin PF, David Ussery D (2004). "CBS Genome Atlas Database: A dynamic storage for bioinformatic results and DNA sequence data". *Bioinformatics* **20**: 3682–3686.
- Zhang CT, Zhang R, Ou HY (2003). "The Z curve database: a graphic representation of genome sequences". *Bioinformatics* **19** (5): 593-599. doi:10.1093/bioinformatics/btg041

External links

- DNA the Double Helix Game ^[152] From the official Nobel Prize web site
- MDDNA: Structural Bioinformatics of DNA ^[50]
- Double Helix 1953–2003 ^[156] National Centre for Biotechnology Education
- DNALive: a web interface to compute DNA physical properties ^[51]. Also allows cross-linking of the results with the UCSC Genome browser and DNA dynamics.
- DiProDB: Dinucleotide Property Database ^[52]. The database is designed to collect and analyse thermodynamic, structural and other dinucleotide properties.
- Further details of mathematical and molecular analysis of DNA structure based on X-ray data ^[53]
- Bessel functions corresponding to Fourier transforms of atomic or molecular helices. ^[54]
- Application of X-ray microscopy in analysis of living hydrated cells ^[19]
- overview of STM/AFM/SNOM principles with educative videos ^[55]

Databases for DNA molecular models and sequences

X-ray diffraction

- NDB ID: UD0017 Database ^[13]
- X-ray Atlas -database ^[56]
- PDB files of coordinates for nucleic acid structures from X-ray diffraction by NA (incl. DNA) crystals ^[57]
- Structure factors downloadable files in CIF format ^[58]

Neutron scattering

- ISIS neutron source: ISIS pulsed neutron source: A world centre for science with neutrons & muons at Harwell, near Oxford, UK. ^[59]

X-ray microscopy

- Application of X-ray microscopy in the analysis of living hydrated cells ^[19]

Electron microscopy

- DNA under electron microscope ^[153]

Genomic and structural databases

- CBS Genome Atlas Database ^[60] — contains examples of base skews.
- The Z curve database of genomes — a 3-dimensional visualization and analysis tool of genomes ^[61].
- DNA and other nucleic acids' molecular models: Coordinate files of nucleic acids molecular structure models in PDB and CIF formats ^[62]

Atomic force microscopy

- How SPM Works ^[63]
 - SPM Image Gallery - AFM STM SEM MFM NSOM and more. ^[64]
-

References

- [1] Franklin, R.E. and Gosling, R.G. recd.6 March 1953. *Acta Cryst.* (1953). 6, 673 The Structure of Sodium Thymonucleate Fibres I. The Influence of Water Content *Acta Cryst.* (1953). and 6, 678 The Structure of Sodium Thymonucleate Fibres II. The Cylindrically Symmetrical Patterson Function.
- [2] Cochran, W., Crick, F.H.C. and Vand V. 1952. The Structure of Synthetic Polypeptides. 1. The Transform of Atoms on a Helix. *Acta Cryst.* 5(5):581-586.
- [3] Crick, F.H.C. 1953a. The Fourier Transform of a Coiled-Coil., *Acta Crystallographica* 6(8-9):685-689.
- [4] Watson, James D. and Francis H.C. Crick. A structure for Deoxyribose Nucleic Acid (<http://www.nature.com/nature/dna50/watsoncrick.pdf>) (PDF). *Nature* 171, 737-738, 25 April 1953.
- [5] Watson, J.D; Crick F.H.C. 1953b. The Structure of DNA., Cold Spring Harbor Symposia on Quantitative Biology 18:123-131.
- [6] Wilkins M.H.F., A.R. Stokes A.R. & Wilson, H.R. (1953). "Molecular Structure of Deoxypentose Nucleic Acids" (<http://www.nature.com/nature/dna50/wilkins.pdf>) (PDF). *Nature* **171** (4356): 738-740. doi:10.1038/171738a0. PMID 13054693. .
- [7] Elson D, Chargaff E (1952). "On the deoxyribonucleic acid content of sea urchin gametes". *Experientia* **8** (4): 143-145.
- [8] Chargaff E, Lipshitz R, Green C (1952). "Composition of the deoxypentose nucleic acids of four genera of sea-urchin". *J Biol Chem* **195** (1): 155-160. PMID 14938364.
- [9] Chargaff E, Lipshitz R, Green C, Hodes ME (1951). "The composition of the deoxyribonucleic acid of salmon sperm". *J Biol Chem* **192** (1): 223-230. PMID 14917668.
- [10] Chargaff E (1951). "Some recent studies on the composition and structure of nucleic acids". *J Cell Physiol Suppl* **38** (Suppl).
- [11] Magasanik B, Vischer E, Doniger R, Elson D, Chargaff E (1950). "The separation and estimation of ribonucleotides in minute quantities". *J Biol Chem* **186** (1): 37-50. PMID 14778802.
- [12] Chargaff E (1950). "Chemical specificity of nucleic acids and mechanism of their enzymatic degradation". *Experientia* **6** (6): 201-209.
- [13] <http://ndbserver.rutgers.edu/atlas/xray/structures/U/ud0017/ud0017.html>
- [14] <http://www.phy.cam.ac.uk/research/bss/molbiophysics.php>
- [15] <http://planetphysics.org/encyclopedia/TheoreticalBiophysics.html>
- [16] <http://commons.wikimedia.org/wiki/Category:DNA>
- [17] Hosemann R., Bagchi R.N., *Direct analysis of diffraction by matter*, North-Holland Publ., Amsterdam – New York, 1962.
- [18] Baianu, I.C. (1978). "X-ray scattering by partially disordered membrane systems.". *Acta Cryst.*, **A34** (5): 751-753. doi:10.1107/S0567739478001540.
- [19] http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=12379938
- [20] Mao, Chengde; Sun, Weiqiong & Seeman, Nadrian C. (16 June 1999). "Designed Two-Dimensional DNA Holliday Junction Arrays Visualized by Atomic Force Microscopy". *Journal of the American Chemical Society* **121** (23): 5437-5443. doi:10.1021/ja9900398. ISSN 0002-7863.
- [21] <http://www.phy.cam.ac.uk/research/bss/molbiophysics.php>
- [22] <http://planetphysics.org/encyclopedia/TheoreticalBiophysics.html>
- [23] (<http://www.jonathanpmiller.com/Karplus.html>)- obtaining dihedral angles from ³J coupling constants
- [24] (http://www.spectroscopynow.com/FCKeditor/UserFiles/File/specNOW/HTML files/General_Karplus_Calculator.htm) Another Javascript-like NMR coupling constant to dihedral
- [25] <http://ndbserver.rutgers.edu/atlas/nmr/index.html>
- [26] <http://ndbserver.rutgers.edu/ftp/NDB/coordinates/na-nmr-mmCIF/>
- [27] <http://ndbserver.rutgers.edu/ftp/NDB/nmr-restraints/>
- [28] Lee, S. C. et al., (2001). One Micrometer Resolution NMR Microscopy. *J. Magn. Res.*, **150**: 207-213.
- [29] Near Infrared Microspectroscopy, Fluorescence Microspectroscopy, Infrared Chemical Imaging and High Resolution Nuclear Magnetic Resonance Analysis of Soybean Seeds, Somatic Embryos and Single Cells., Baianu, I.C. et al. 2004., In *Oil Extraction and Analysis.*, D. Luthria, Editor pp.241-273, AOCS Press., Champaign, IL.
- [30] Single Cancer Cell Detection by Near Infrared Microspectroscopy, Infrared Chemical Imaging and Fluorescence Microspectroscopy. 2004.I. C. Baianu, D. Costescu, N. E. Hofmann, S. S. Korban and et al., q-bio/0407006 (July 2004) (<http://arxiv.org/abs/q-bio/0407006>)
- [31] Raghavachari, R., Editor. 2001. *Near-Infrared Applications in Biotechnology*, Marcel-Dekker, New York, NY.
- [32] [http://www.imaging.net/chemical-imaging/Chemical imaging](http://www.imaging.net/chemical-imaging/Chemical%20imaging)
- [33] http://www.malvern.com/LabEng/products/sdi/bibliography/sdi_bibliography.htm E. N. Lewis, E. Lee and L. H. Kidder, Combining Imaging and Spectroscopy: Solving Problems with Near-Infrared Chemical Imaging. *Microscopy Today*, Volume 12, No. 6, 11/2004.
- [34] D.S. Mantus and G. H. Morrison. 1991. Chemical imaging in biology and medicine using ion microscopy., *Microchimica Acta*, **104**, (1-6) January 1991, doi: 10.1007/BF01245536
- [35] J. Dubois, G. Sando, E. N. Lewis, Near-Infrared Chemical Imaging, A Valuable Tool for the Pharmaceutical Industry, G.I.T. Laboratory Journal Europe, No.1-2, 2007.
- [36] Applications of Novel Techniques to Health Foods, Medical and Agricultural Biotechnology.(June 2004).,I. C. Baianu, P. R. Lozano, V. I. Prisecaru and H. C. Lin q-bio/0406047 (<http://arxiv.org/abs/q-bio/0406047>)
- [37] Chemical Imaging Without Dyeing (<http://witec.de/en/download/Raman/ImagingMicroscopy04.pdf>)

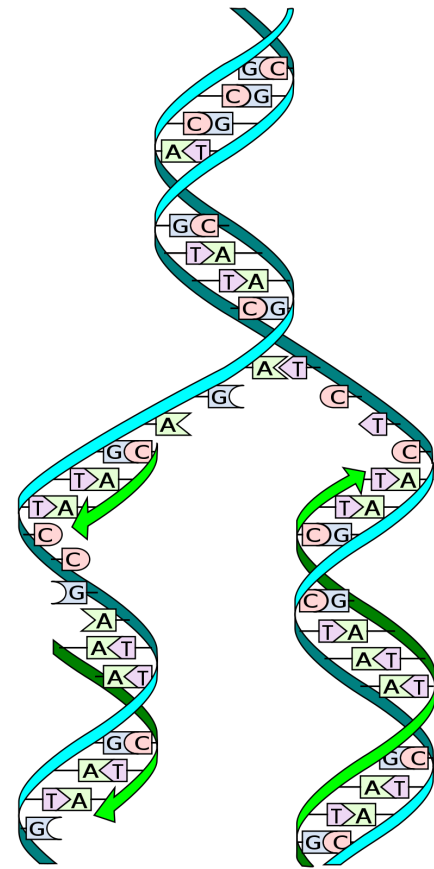
- [38] C.L. Evans and X.S. Xie.2008. Coherent Anti-Stokes Raman Scattering Microscopy: Chemical Imaging for Biology and Medicine., doi:10.1146/annurev.anchem.1.031207.112754 *Annual Review of Analytical Chemistry*, **1**: 883-909.
- [39] Eigen, M., Rigler, R. Sorting single molecules: application to diagnostics and evolutionary biotechnology,(1994) *Proc. Natl. Acad. Sci. USA*, **91**,5740-5747.
- [40] Rigler, M. Fluorescence correlations, single molecule detection and large number screening. Applications in biotechnology,(1995) *J. Biotechnol.*, **41**,177-186.
- [41] Rigler R. and Widengren J. (1990). Ultrasensitive detection of single molecules by fluorescence correlation spectroscopy, *BioScience* (Ed. Klinge & Owman) p.180.
- [42] Oehlenschläger F., Schwille P. and Eigen M. (1996). Detection of HIV-1 RNA by nucleic acid sequence-based amplification combined with fluorescence correlation spectroscopy, *Proc. Natl. Acad. Sci. USA* **93**:1281.
- [43] Bagatolli, L.A., and Gratton, E. (2000). Two-photon fluorescence microscopy of coexisting lipid domains in giant unilamellar vesicles of binary phospholipid mixtures. *Biophys J.*, **78**:290-305.
- [44] Schwille, P., Haupts, U., Maiti, S., and Webb. W.(1999). Molecular dynamics in living cells observed by fluorescence correlation spectroscopy with one- and two-photon excitation. *Biophysical Journal*, **77**(10):2251-2265.
- [45] FRET description (<http://dwb.unl.edu/Teacher/NSF/C08/C08Links/pp99.cryst.bbk.ac.uk/projects/gmocz/fret.htm>)
- [46] doi:10.1016/S0959-440X(00)00190-1 ([http://dx.doi.org/10.1016/S0959-440X\(00\)00190-1](http://dx.doi.org/10.1016/S0959-440X(00)00190-1))Recent advances in FRET: distance determination in protein–DNA complexes. *Current Opinion in Structural Biology* **2001**, *11*(2), 201-207
- [47] <http://www.fretimaging.org/mcnamaraintro.html> FRET imaging introduction
- [48] <http://peds.oxfordjournals.org/cgi/content/abstract/1/4/295>
- [49] <http://www.sciencemag.org/cgi/content/abstract/sci;302/5649/1380>
- [50] <http://humphry.chem.wesleyan.edu:8080/MDDNA/>
- [51] <http://mmb.pcb.ub.es/DNAlive>
- [52] <http://diprodb.fli-leibniz.de>
- [53] <http://planetphysics.org/encyclopedia/BesselFunctionsApplicationsToDiffractionByHelicalStructures.html>
- [54] <http://planetphysics.org/?op=getobj&from=objects&name=BesselFunctionsAndTheirApplicationsToDiffractionByHelicalStructures>
- [55] <http://www.ntmdt.ru/SPM-Techniques/Principles/>
- [56] <http://ndbserver.rutgers.edu/atlas/xray/index.html>
- [57] <http://ndbserver.rutgers.edu/ftp/NDB/coordinates/na-biol/>
- [58] <http://ndbserver.rutgers.edu/ftp/NDB/structure-factors/>
- [59] <http://www.isis.rl.ac.uk/>
- [60] <http://www.cbs.dtu.dk/services/GenomeAtlas/>
- [61] <http://tubic.tju.edu.cn/zcurve/>
- [62] <http://ndbserver.rutgers.edu/ftp/NDB/models/>
- [63] http://www.parkafm.com/New_html/resources/01general.php
- [64] <http://www.rhk-tech.com/results/showcase.php>

DNA replication

DNA replication, the basis for biological inheritance, is a fundamental process occurring in all living organisms to copy their DNA. This process is "replication" in that each strand of the original double-stranded DNA molecule serves as template for the reproduction of the complementary strand. Hence, following DNA replication, two identical DNA molecules have been produced from a single double-stranded DNA molecule. Cellular proofreading and error checking mechanisms ensure near perfect fidelity for DNA replication.^{[1] [2]}

In a cell, DNA replication begins at specific locations in the genome, called "origins".^[3] Unwinding of DNA at the origin, and synthesis of new strands, forms a replication fork. In addition to DNA polymerase, the enzyme that synthesizes the new DNA by adding nucleotides matched to the template strand, a number of other proteins are associated with the fork and assist in the initiation and continuation of DNA synthesis.

DNA replication can also be performed *in vitro* (outside a cell). DNA polymerases, isolated from cells, and artificial DNA primers are used to initiate DNA synthesis at known sequences in a template molecule. The polymerase chain reaction (PCR), a common laboratory technique, employs such artificial synthesis in a cyclic manner to amplify a specific target DNA fragment from a pool of DNA.



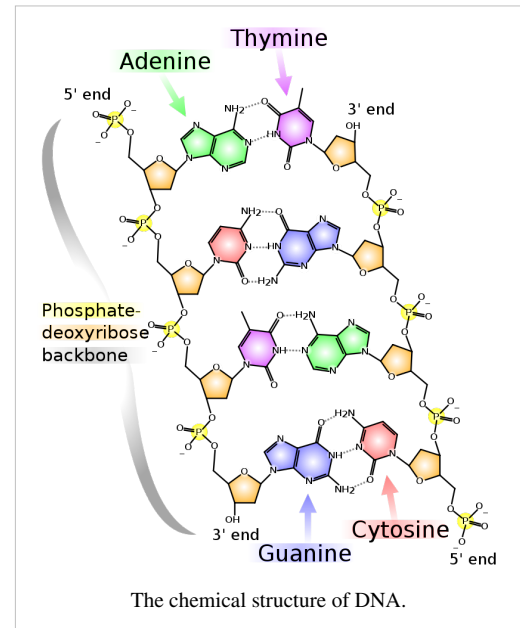
DNA replication. The double helix is unwound and each strand acts as a template. Bases are matched to synthesize the new partner strands.

DNA structure

DNA usually exists as a double-stranded structure, with both strands coiled together to form the characteristic double-helix. Each single strand of DNA is a chain of four types of nucleotides: adenine, cytosine, guanine, and thymine. A nucleotide is a mono-, di- or triphosphate deoxyribonucleoside; that is, a deoxyribose sugar is attached to one, two or three phosphates. Chemical interaction of these nucleotides forms phosphodiester linkages, creating the phosphate-deoxyribose backbone of the DNA double helix with the bases pointing inward. Nucleotides (bases) are matched between strands through hydrogen bonds to form base pairs. Adenine pairs with thymine and cytosine pairs with guanine.

DNA strands have a directionality, and the different ends of a single strand are called the "3' (three-prime) end" and the "5' (five-prime) end." These terms refer to the carbon atom in deoxyribose to which the next phosphate in the chain attaches. In addition to being complementary, the two strands of DNA are antiparallel: they are oriented in opposite directions. This directionality has consequences in DNA synthesis, because DNA polymerase can only synthesize DNA in one direction by adding nucleotides to the 3' end of a DNA strand.

The pairing of bases in DNA through hydrogen bonding means that the information contained within each strand is redundant. The nucleotides on a single strand can be used to reconstruct nucleotides on a newly synthesized partner strand.^[4]

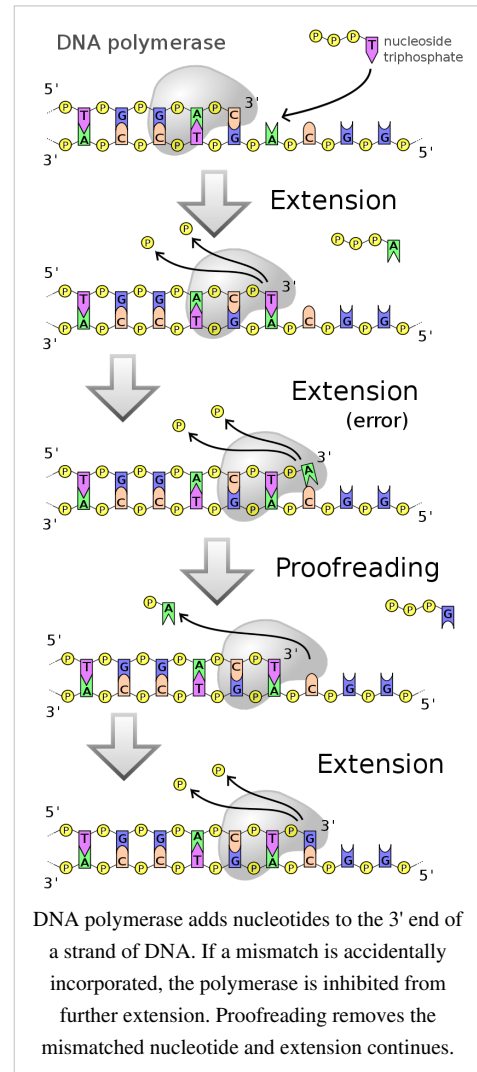


DNA polymerase

DNA polymerases are a family of enzymes that carry out all forms of DNA replication.^[5] A DNA polymerase can only extend an existing DNA strand paired with a template strand; it cannot begin the synthesis of a new strand. To begin synthesis of a new strand, a short fragment of DNA or RNA, called a primer, must be created and paired with the template strand before DNA polymerase can synthesize new DNA.

Once a primer pairs with DNA to be replicated, DNA polymerase synthesizes a new strand of DNA by extending the 3' end of an existing nucleotide chain, adding new nucleotides matched to the template strand one at a time via the creation of phosphodiester bonds. The energy for this process of DNA polymerization comes from two of the three total phosphates attached to each unincorporated base. (Free bases with their attached phosphate groups are called nucleoside triphosphates.) When a nucleotide is being added to a growing DNA strand, two of the phosphates are removed and the energy produced creates a phosphodiester (chemical) bond that attaches the remaining phosphate to the growing chain. The energetics of this process also help explain the directionality of synthesis - if DNA were synthesized in the 3' to 5' direction, the energy for the process would come from the 5' end of the growing strand rather than from free nucleotides.

DNA polymerases are generally extremely accurate, making less than one error for every 10^7 nucleotides added.^[6] Even so, some DNA polymerases also have proofreading ability; they can remove nucleotides from the end of a strand in order to correct mismatched bases. If the 5' nucleotide needs to be removed during proofreading, the triphosphate end is lost. Hence, the energy source that usually provides energy to add a new nucleotide is also lost.



DNA replication within the cell

Origins of replication

For a cell to divide, it must first replicate its DNA.^[7] This process is initiated at particular points within the DNA, known as "origins", which are targeted by proteins that separate the two strands and initiate DNA synthesis.^[3] Origins contain DNA sequences recognized by replication initiator proteins (eg. *dnaA* in *E. coli* and the Origin Recognition Complex in yeast).^[8] These initiator proteins recruit other proteins to separate the two strands and initiate replication forks.

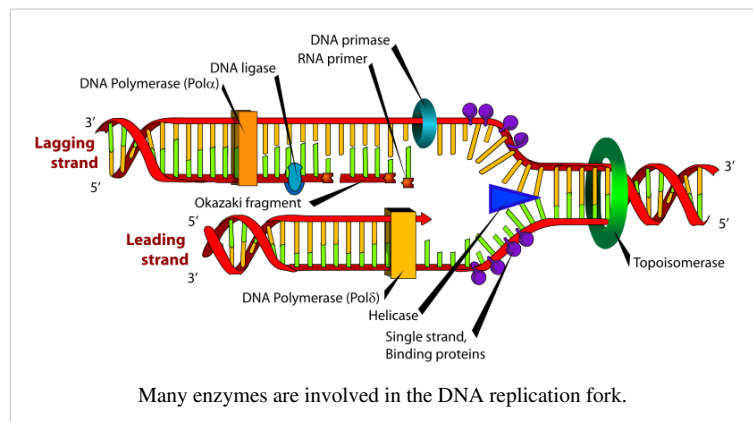
Initiator proteins recruit other proteins to separate the DNA strands at the origin, forming a bubble. Origins tend to be "AT-rich" (rich in adenine and thymine bases) to assist this process, because A-T base pairs have two hydrogen bonds (rather than the three formed in a C-G pair)—strands rich in these nucleotides are generally easier to separate due to the positive relationship between the number of hydrogen bonds and the difficulty of breaking these bonds.^[9] Once strands are separated, RNA primers are created on the template strands. More specifically, the leading strand receives one RNA primer per active origin of replication while the lagging strand receives several; these several fragments of RNA primers found on the lagging strand of DNA are called Okazaki fragments, named after their

discoverer. DNA polymerase extends the leading strand in one continuous motion and the lagging strand in a discontinuous motion (due to the Okazaki fragments). RNase removes the RNA fragments used to initiate replication by DNA Polymerase, and another DNA Polymerase enters to fill the gaps. When this is complete, a single nick on the leading strand and several nicks on the lagging strand can be found. Ligase works to fill these nicks in, thus completing the newly replicated DNA molecule.

As DNA synthesis continues, the original DNA strands continue to unwind on each side of the bubble, forming 2 replication forks. In bacteria, which have a single origin of replication on their circular chromosome, this process eventually creates a "theta structure" (resembling the Greek letter theta: θ). In contrast, eukaryotes have longer linear chromosomes and initiate replication at multiple origins within these.

The replication fork

When replicating, the original DNA splits in two, forming two "prongs" which resemble a fork (hence the name "replication fork"). DNA has a ladder-like structure; imagine a ladder broken in half vertically, along the steps. Each half of the ladder now requires a new half to match it. Because DNA polymerase can only synthesize a new DNA strand in a 5' to 3' manner, the process of replication goes differently for the two strands comprising the DNA double helix.



Leading strand

The leading strand is that strand of the DNA double helix that is oriented in a 5' to 3' manner.

On the leading strand, a polymerase "reads" the DNA and adds nucleotides to it continuously. This polymerase is DNA polymerase III (DNA Pol III) in prokaryotes and presumably Pol ϵ ^[10] ^[11] in eukaryotes.

Lagging strand

The lagging strand is that strand of the DNA double helix that is orientated in a 3' to 5' manner. Because of its orientation, opposite to the working orientation of DNA polymerase III which is in a 3' to 5' manner, replication of the lagging strand is more complicated than that of the leading strand.

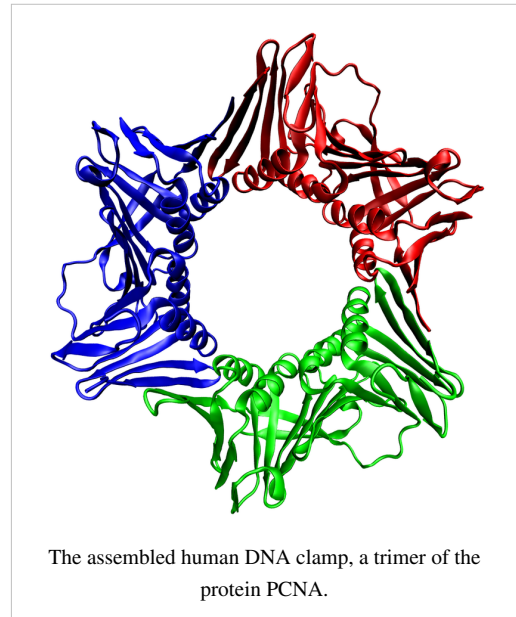
On the lagging strand, primase "reads" the DNA and adds RNA to it in short, separated segments. In eukaryotes, primase is intrinsic to Pol α .^[12] DNA polymerase III or Pol δ lengthens the primed segments, forming Okazaki fragments. Primer removal in eukaryotes is also performed by Pol δ .^[13] In prokaryotes, DNA polymerase I "reads" the fragments, removes the RNA using its flap endonuclease domain, and replaces the RNA nucleotides with DNA nucleotides (this is necessary because RNA and DNA use slightly different kinds of nucleotides). DNA ligase joins the fragments together.

Dynamics at the replication fork

As helicase unwinds DNA at the replication fork, the DNA ahead is forced to rotate. This process results in a build-up of twists in the DNA ahead.^[14] This build-up would form a resistance that would eventually halt the progress of the replication fork. DNA topoisomerases are enzymes that solve these physical problems in the coiling of DNA. Topoisomerase I cuts a single backbone on the DNA, enabling the strands to swivel around each other to remove the build-up of twists. Topoisomerase II cuts both backbones, enabling one double-stranded DNA to pass through another, thereby removing knots and entanglements that can form within and between DNA molecules.

Bare single-stranded DNA has a tendency to fold back upon itself and form secondary structures; these structures can interfere with the movement of DNA polymerase. To prevent this, single-strand binding proteins bind to the DNA until a second strand is synthesized, preventing secondary structure formation.^[15]

Clamp proteins form a sliding clamp around DNA, helping the DNA polymerase maintain contact with its template and thereby assisting with processivity. The inner face of the clamp enables DNA to be threaded through it. Once the polymerase reaches the end of the template or detects double stranded DNA, the sliding clamp undergoes a conformational change which releases the DNA polymerase. Clamp-loading proteins are used to initially load the clamp, recognizing the junction between template and RNA primers.



Regulation of replication

Eukaryotes

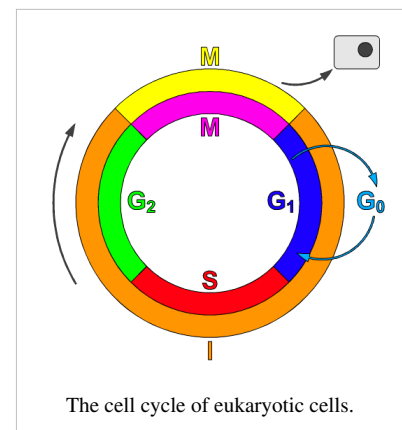
Within eukaryotes, DNA replication is controlled within the context of the cell cycle. As the cell grows and divides, it progresses through stages in the cell cycle; DNA replication occurs during the S phase (Synthesis phase). The progress of the eukaryotic cell through the cycle is controlled by cell cycle checkpoints. Progression through checkpoints is controlled through complex interactions between various proteins, including cyclins and cyclin-dependent kinases.^[16]

The G1/S checkpoint (or restriction checkpoint) regulates whether eukaryotic cells enter the process of DNA replication and subsequent division. Cells which do not proceed through this checkpoint are quiescent in the "G0" stage and do not replicate their DNA.

Replication of chloroplast and mitochondrial genomes occurs independent of the cell cycle, through the process of D-loop replication.

Bacteria

Most bacteria do not go through a well-defined cell cycle and instead continuously copy their DNA; during rapid growth this can result in multiple rounds of replication occurring concurrently.^[17] Within *E. coli*, the most well-characterized bacteria, regulation of DNA replication can be achieved through several mechanisms, including: the hemimethylation and sequestering of the origin sequence, the ratio of ATP to ADP, and the levels of protein DnaA. These all control the process of initiator proteins binding to the origin sequences.



Because *E. coli* methylates GATC DNA sequences, DNA synthesis results in hemimethylated sequences. This hemimethylated DNA is recognized by a protein (SeqA) which binds and sequesters the origin sequence; in addition, dnaA (required for initiation of replication) binds less well to hemimethylated DNA. As a result, newly replicated origins are prevented from immediately initiating another round of DNA replication.^[18]

ATP builds up when the cell is in a rich medium, triggering DNA replication once the cell has reached a specific size. ATP competes with ADP to bind to DnaA, and the DnaA-ATP complex is able to initiate replication. A certain number of DnaA proteins are also required for DNA replication — each time the origin is copied the number of binding sites for DnaA doubles, requiring the synthesis of more DnaA to enable another initiation of replication.

Termination of replication

Because bacteria have circular chromosomes, termination of replication occurs when the two replication forks meet each other on the opposite end of the parental chromosome. *E. coli* regulate this process through the use of termination sequences which, when bound by the Tus protein, enable only one direction of replication fork to pass through. As a result, the replication forks are constrained to always meet within the termination region of the chromosome.^[19]

Eukaryotes initiate DNA replication at multiple points in the chromosome, so replication forks meet and terminate at many points in the chromosome; these are not known to be regulated in any particular manner. Because eukaryotes have linear chromosomes, DNA replication often fails to synthesize to the very end of the chromosomes (telomeres), resulting in telomere shortening. This is a normal process in somatic cells — cells are only able to divide a certain number of times before the DNA loss prevents further division. (This is known as the Hayflick limit.) Within the germ cell line, which passes DNA to the next generation, telomerase extends the repetitive sequences of the telomere region to prevent degradation. Telomerase can become mistakenly active in somatic cells, sometimes leading to cancer formation.

Polymerase chain reaction

Researchers commonly replicate DNA *in vitro* using the polymerase chain reaction (PCR). PCR uses a pair of primers to span a target region in template DNA, and then polymerizes partner strands in each direction from these primers using a thermostable DNA polymerase. Repeating this process through multiple cycles produces amplification of the targeted DNA region. At the start of each cycle, the mixture of template and primers is heated, separating the newly synthesized molecule and template. Then, as the mixture cools, both of these become templates for annealing of new primers, and the polymerase extends from these. As a result, the number of copies of the target region doubles each round, increasing exponentially.^[20]

References

- [1] Berg JM, Tymoczko JL, Strxcvyer L, Clarke ND (2002). *Biochemistry*. W.H. Freeman and Company. ISBN 0-7167-3051-0. Chapter 27: DNA Replication, Recombination, and Repair (<http://www.ncbi.nlm.nih.gov/books/bv.fc.kgi?rid=stryer.chapter.3740>)
- [2] Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P (2002). *Molecular Biology of the Cell*. Garland Science. ISBN 0-8153-3218-1. Chapter 5: DNA Replication, Repair, and Recombination (<http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=mboc4.chapter.747>)
- [3] Berg JM, Tymoczko JL, Stryer L, Clarke ND (2002). *Biochemistry*. W.H. Freeman and Company. ISBN 0-7167-3051-0. Chapter 27, Section 4: DNA Replication of Both Strands Proceeds Rapidly from Specific Start Sites (<http://www.ncbi.nlm.nih.gov/bofghoks/bv.fcvbngi?rid=stryer.section.3794>)
- [4] Alberts, B., et.al., *Molecular Biology of the Cell*, Garland Science, 4th ed., 2002, pp. 238-240 ISBN 0-8153-3218-1
- [5] Berg JM, Tymoczko JL, Stryer L, Clarke ND (2002). *Biochemistry*. W.H. Freeman and Company. ISBN 0-7167-3051-0. Chapter 27, Section 2: DNA Polymerases Require a Template and a Primer (<http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=stryer.section.3769>)
- [6] McCulloch SD, Kunkel TA (January 2008). "The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases". *Cell Research* **18** (1): 148–61. doi:10.1038/cr.2008.4. PMID 18166979.
- [7] Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P (2002). *Molecular Biology of the Cell*. Garland Science. ISBN 0-8153-3218-1. Chapter 5: DNA Replication Mechanisms (<http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=mboc4.section.754>)

- [8] Weigel C, Schmidt A, Rückert B, Lurz R, Messer W (November 1997). "DnaA protein binding to individual DnaA boxes in the Escherichia coli replication origin, oriC" (<http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=1170261>). *The EMBO Journal* **16** (21): 6574–83. doi:10.1093/emboj/16.21.6574. PMID 9351837. PMC 1170261.
- [9] Lodish H, Berk A, Zipursky LS, Matsudaira P, Baltimore D, Darnell J (2000). *Molecular Cell Biology*. W. H. Freeman and Company. ISBN 0-7167-3136-3. 12.1. General Features of Chromosomal Replication: Three Common Features of Replication Origins (<http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=mcb.section.3163#3179>)
- [10] Pursell, Z.F. et al. (2007). "Yeast DNA Polymerase ϵ Participates in Leading-Strand DNA Replication" (<http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=2233713>). *Science* **317** (5834): 127. doi:10.1126/science.1144067. PMID 17615360. PMC 2233713.
- [11] Scott D McCulloch; Thomas A Kunkel (01/2008). "The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases". *Cell Research* **18** (1): 148. doi:10.1038/cr.2008.4. PMID 18166979.
- [12] Elizabeth R. Barry; Stephen D. Bell (12/2006). "DNA Replication in the Archaea" (<http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=1698513>). *Microbiology and Molecular Biology Reviews* **70** (4): 876–887. doi:10.1128/MMBR.00029-06. PMID 17158702. PMC 1698513.
- [13] Distinguishing the pathways of primer removal during Eukaryotic Okazaki fragment maturation (<http://hdl.handle.net/1802/6537>) Contributor Author Rossi, Marie Louise. Date Accessioned: 2009-02-23T17:05:09Z. Date Available: 2009-02-23T17:05:09Z. Date Issued: 2009-02-23T17:05:09Z. Identifier Uri: <http://hdl.handle.net/1802/6537>. Description: Dr. Robert A. Bambara, Faculty Advisor. Thesis (PhD) - School of Medicine and Dentistry, University of Rochester. UR only until January 2010. UR only until January 2010.
- [14] Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P (2002). *Molecular Biology of the Cell*. Garland Science. ISBN 0-8153-3218-1. DNA Replication Mechanisms: DNA Topoisomerases Prevent DNA Tangling During Replication (<http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=mboc4.section.754#787>)
- [15] Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P (2002). *Molecular Biology of the Cell*. Garland Science. ISBN 0-8153-3218-1. DNA Replication Mechanisms: Special Proteins Help to Open Up the DNA Double Helix in Front of the Replication Fork (<http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=mboc4.section.754#774>)
- [16] Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P (2002). *Molecular Biology of the Cell*. Garland Science. ISBN 0-8153-3218-1. Intracellular Control of Cell-Cycle Events: S-Phase Cyclin-Cdk Complexes (S-Cdks) Initiate DNA Replication Once Per Cycle (<http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=mboc4.section.3214#3215>)
- [17] Tobiason DM, Seifert HS (2006). "The Obligate Human Pathogen, *Neisseria gonorrhoeae*, Is Polyploid" (<http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=1470461>). *PLoS Biology* **4** (6): e185. doi:10.1371/journal.pbio.0040185. PMID 16719561. PMC 1470461.
- [18] Slater S, Wold S, Lu M, Boye E, Skarstad K, Kleckner N (September 1995). "E. coli SeqA protein binds oriC in two different methyl-modulated reactions appropriate to its roles in DNA replication initiation and origin sequestration". *Cell* **82** (6): 927–36. doi:10.1016/0092-8674(95)90272-4. PMID 7553853.
- [19] TA Brown (2002). *Genomes*. BIOS Scientific Publishers Ltd. ISBN 1 85996 228 9. 13.2.3. Termination of replication (<http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=genomes.section.8121#8156>)
- [20] Saiki, RK; Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA (1988). "Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase" (<http://sunsite.berkeley.edu/cgi-bin/ebind2html/pcr/009>). *Science* **239** (4839): 487–91. doi:10.1126/science.2448875. PMID 2448875. .
-

DNA repair

DNA repair refers to a collection of processes by which a cell identifies and corrects damage to the DNA molecules that encode its genome. In human cells, both normal metabolic activities and environmental factors such as UV light and radiation can cause DNA damage, resulting in as many as 1 million individual molecular lesions per cell per day.^[1] Many of these lesions cause structural damage to the DNA molecule and can alter or eliminate the cell's ability to transcribe the gene that the affected DNA encodes. Other lesions induce potentially harmful mutations in the cell's genome, which affect the survival of its daughter cells after it undergoes mitosis. Consequently, the DNA repair process is constantly active as it responds to damage in the DNA structure. When normal repair processes fail, and when cellular apoptosis does not occur, irreparable DNA damage may occur, including double-strand breaks and DNA crosslinkages.^{[2] [3]}

The rate of DNA repair is dependent on many factors, including the cell type, the age of the cell, and the extracellular environment. A cell that has accumulated a large amount of DNA damage, or one that no longer effectively repairs damage incurred to its DNA, can enter one of three possible states:

1. an irreversible state of dormancy, known as senescence
2. cell suicide, also known as apoptosis or programmed cell death
3. unregulated cell division, which can lead to the formation of a tumor that is cancerous

The DNA repair ability of a cell is vital to the integrity of its genome and thus to its normal functioning and that of the organism. Many genes that were initially shown to influence life span have turned out to be involved in DNA damage repair and protection.^[4] Failure to correct molecular lesions in cells that form gametes can introduce mutations into the genomes of the offspring and thus influence the rate of evolution.

DNA damage

DNA damage, due to environmental factors and normal metabolic processes inside the cell, occurs at a rate of 1,000 to 1,000,000 molecular lesions per cell per day.^[1] While this constitutes only 0.000165% of the human genome's approximately 6 billion bases (3 billion base pairs), unrepaired lesions in critical genes (such as tumor suppressor genes) can impede a cell's ability to carry out its function and appreciably increase the likelihood of tumor formation.

The vast majority of DNA damage affects the primary structure of the double helix; that is, the bases themselves are chemically modified. These modifications can in turn disrupt the molecules' regular helical structure by introducing non-native chemical bonds or bulky adducts that do not fit in the standard double helix. Unlike proteins and RNA, DNA usually lacks tertiary structure and therefore damage or disturbance does not occur at that level. DNA is, however, supercoiled and wound around "packaging" proteins called histones (in eukaryotes), and both superstructures are vulnerable to the effects of DNA damage.



DNA damage resulting in multiple broken chromosomes

Sources of damage

DNA damage can be subdivided into two main types:

1. endogenous damage such as attack by reactive oxygen species produced from normal metabolic byproducts (spontaneous mutation), especially the process of oxidative deamination;
 1. also includes replication errors
2. exogenous damage caused by external agents such as
 1. ultraviolet [UV 200-300nm] radiation from the sun
 2. other radiation frequencies, including x-rays and gamma rays
 3. hydrolysis or thermal disruption
 4. certain plant toxins
 5. human-made mutagenic chemicals, especially aromatic compounds that act as DNA intercalating agents
 6. cancer chemotherapy and radiotherapy
 7. viruses ^[5]

The replication of damaged DNA before cell division can lead to the incorporation of wrong bases opposite damaged ones. Daughter cells that inherit these wrong bases carry mutations from which the original DNA sequence is unrecoverable (except in the rare case of a back mutation, for example, through gene conversion).

Types of damage

There are five main types of damage to DNA due to endogenous cellular processes:

1. *oxidation* of bases [e.g. 8-oxo-7,8-dihydroguanine (8-oxoG)] and generation of DNA strand interruptions from reactive oxygen species,
2. *alkylation* of bases (usually methylation), such as formation of 7-methylguanine, 1-methyladenine, 6-O-Methylguanine
3. *hydrolysis* of bases, such as deamination, depurination and depyrimidination.
4. "bulky adduct formation" (i.e. benzo[a]pyrene diol epoxide-dG adduct)
5. *mismatch* of bases, due to errors in DNA replication, in which the wrong DNA base is stitched into place in a newly forming DNA strand, or a DNA base is skipped over or mistakenly inserted.

Damage caused by exogenous agents comes in many forms. Some examples are:

1. *UV-B light* causes crosslinking between adjacent cytosine and thymine bases creating *pyrimidine dimers*. This is called direct DNA damage.
2. *UV-A light* creates mostly free radicals. The damage caused by free radicals is called indirect DNA damage.
3. *Ionizing radiation* such as that created by radioactive decay or in *cosmic rays* causes breaks in DNA strands.
4. *Thermal disruption* at elevated temperature increases the rate of depurination (loss of purine bases from the DNA backbone) and single strand breaks. For example, hydrolytic depurination is seen in the thermophilic bacteria, which grow in hot springs at 40-80 °C^[6].^[7] The rate of depurination (300 purine residues per genome per generation) is too high in these species to be repaired by normal repair machinery, hence a possibility of an adaptive response cannot be ruled out.
5. *Industrial chemicals* such as vinyl chloride and hydrogen peroxide, and environmental chemicals such as polycyclic hydrocarbons found in smoke, soot and tar create a huge diversity of DNA adducts- ethenobases, oxidized bases, alkylated phosphotriesters and Crosslinking of DNA just to name a few.

UV damage, alkylation/methylation, X-ray damage and oxidative damage are examples of induced damage. Spontaneous damage can include the loss of a base, deamination, sugar ring puckering and tautomeric shift.^[8]

Nuclear versus mitochondrial DNA damage

In human cells, and eukaryotic cells in general, DNA is found in two cellular locations - inside the nucleus and inside the mitochondria. Nuclear DNA (nDNA) exists as chromatin during non-replicative stages of the cell cycle and is condensed into aggregate structures known as chromosomes during cell division. In either state the DNA is highly compacted and wound up around bead-like proteins called histones. Whenever a cell needs to express the genetic information encoded in its nDNA the required chromosomal region is unravelled, genes located therein are expressed, and then the region is condensed back to its resting conformation. Mitochondrial DNA (mtDNA) is located inside mitochondria organelles, exists in multiple copies, and is also tightly associated with a number of proteins to form a complex known as the nucleoid. Inside mitochondria, reactive oxygen species (ROS), or free radicals, byproducts of the constant production of adenosine triphosphate (ATP) via oxidative phosphorylation, create a highly oxidative environment that is known to damage mtDNA. A critical enzyme in counteracting the toxicity of these species is superoxide dismutase, which is present in both the mitochondria and cytoplasm of eukaryotic cells.

Senescence and apoptosis

Senescence, an irreversible state in which the cell no longer divides, is a protective response to the shortening of the chromosome ends. The telomeres are long regions of repetitive noncoding DNA that cap chromosomes and undergo partial degradation each time a cell undergoes division (see Hayflick limit).^[9] In contrast, quiescence is a reversible state of cellular dormancy that is unrelated to genome damage (see cell cycle). Senescence in cells may serve as a functional alternative to apoptosis in cases where the physical presence of a cell for spatial reasons is required by the organism,^[10] which serves as a "last resort" mechanism to prevent a cell with damaged DNA from replicating inappropriately in the absence of pro-growth cellular signaling. Unregulated cell division can lead to the formation of a tumor (see cancer), which is potentially lethal to an organism. Therefore the induction of senescence and apoptosis is considered to be part of a strategy of protection against cancer.^[11]

DNA damage and mutation

It is important to distinguish between DNA damage and mutation, the two major types of error in DNA. DNA damages and mutation are fundamentally different. Damages are physical abnormalities in the DNA, such as single and double strand breaks, 8-hydroxydeoxyguanosine residues and polycyclic aromatic hydrocarbon adducts. DNA damages can be recognized by enzymes, and thus they can be correctly repaired if redundant information, such as the undamaged sequence in the complementary DNA strand or in a homologous chromosome, is available for copying. If a cell retains DNA damage, transcription of a gene can be prevented and thus translation into a protein will also be blocked. Replication may also be blocked and/or the cell may die.

In contrast to DNA damage, a mutation is a change in the base sequence of the DNA. A mutation cannot be recognized by enzymes once the base change is present in both DNA strands, and thus a mutation cannot be repaired. At the cellular level, mutations can cause alterations in protein function and regulation. Mutations are replicated when the cell replicates. In a population of cells, mutant cells will increase or decrease in frequency according to the effects of the mutation on the ability of the cell to survive and reproduce. Although distinctly different from each other, DNA damages and mutations are related because DNA damages often cause errors of DNA synthesis during replication or repair and these errors are a major source of mutation.

Given these properties of DNA damage and mutation, it can be seen that DNA damages are a special problem in non-dividing or slowly dividing cells, where unrepaired damages will tend to accumulate over time. On the other hand, in rapidly dividing cells, unrepaired DNA damages that do not kill the cell by blocking replication will tend to cause replication errors and thus mutation. The great majority of mutations that are not neutral in their effect are deleterious to a cell's survival. Thus, in a population of cells comprising a tissue with replicating cells, mutant cells will tend to be lost. However infrequent mutations that provide a survival advantage will tend to clonally expand at

the expense of neighboring cells in the tissue. This advantage to the cell is disadvantageous to the whole organism, because such mutant cells can give rise to cancer. Thus DNA damages in frequently dividing cells, because they give rise to mutations, are a prominent cause of cancer. In contrast, DNA damages in infrequently dividing cells are likely a prominent cause of aging.^[12]

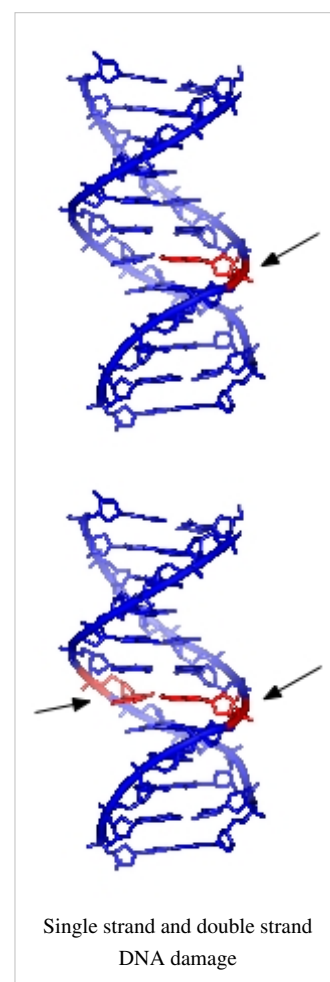
DNA repair mechanisms

Cells cannot function if DNA damage corrupts the integrity and accessibility of essential information in the genome (but cells remain superficially functional when so-called "non-essential" genes are missing or damaged). Depending on the type of damage inflicted on the DNA's double helical structure, a variety of repair strategies have evolved to restore lost information. If possible, cells use the unmodified complementary strand of the DNA or the sister chromatid as a template to recover the original information. Without access to a template, cells use an error-prone recovery mechanism known as translesion synthesis as a last resort.

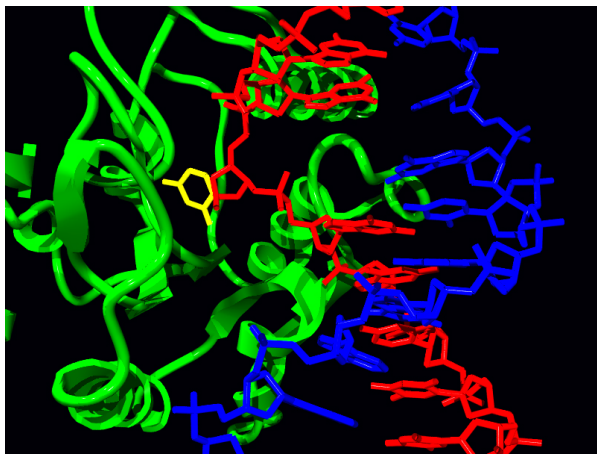
Damage to DNA alters the spatial configuration of the helix and such alterations can be detected by the cell. Once damage is localized, specific DNA repair molecules bind at or near the site of damage, inducing other molecules to bind and form a complex that enables the actual repair to take place. The types of molecules involved and the mechanism of repair that is mobilized depend on the type of damage that has occurred and the phase of the cell cycle that the cell is in.

Direct reversal

Cells are known to eliminate three types of damage to their DNA by chemically reversing it. These mechanisms do not require a template, since the types of damage they counteract can only occur in one of the four bases. Such direct reversal mechanisms are specific to the type of damage incurred and do not involve breakage of the phosphodiester backbone. The formation of pyrimidine dimers upon irradiation with UV light results in an abnormal covalent bond between adjacent pyrimidine bases. The photoreactivation process directly reverses this damage by the action of the enzyme photolyase, whose activation is obligately dependent on energy absorbed from blue/UV light (300–500 nm wavelength) to promote catalysis.^[13] Another type of damage, methylation of guanine bases, is directly reversed by the protein methyl guanine methyl transferase (MGMT), the bacterial equivalent of which is called ogt. This is an expensive process because each MGMT molecule can only be used once; that is, the reaction is stoichiometric rather than catalytic.^[14] A generalized response to methylating agents in bacteria is known as the adaptive response and confers a level of resistance to alkylating agents upon sustained exposure by upregulation of alkylation repair enzymes.^[15] The third type of DNA damage reversed by cells is certain methylation of the bases cytosine and adenine.



Single strand damage



Structure of the base-excision repair enzyme uracil-DNA glycosylase. The uracil residue is shown in yellow.

When only one of the two strands of a double helix has a defect, the other strand can be used as a template to guide the correction of the damaged strand. In order to repair damage to one of the two paired molecules of DNA, there exist a number of excision repair mechanisms that remove the damaged nucleotide and replace it with an undamaged nucleotide complementary to that found in the undamaged DNA strand.^[14]

1. Base excision repair (BER), which repairs damage to a single base caused by oxidation, alkylation, hydrolysis, or deamination. The damaged base is removed by a DNA glycosylase. The "missing tooth" is then recognised by an enzyme called AP endonuclease, which cuts the Phosphodiester bond. The missing part is then resynthesized by a DNA polymerase, and a DNA ligase performs the final nick-sealing step.
2. Nucleotide excision repair (NER), which recognizes bulky, helix-distorting lesions such as pyrimidine dimers and 6,4 photoproducts. A specialized form of NER known as transcription-coupled repair deploys NER enzymes to genes that are being actively transcribed.
3. Mismatch repair (MMR), which corrects errors of DNA replication and recombination that result in mispaired (but undamaged) nucleotides.

Double-strand breaks

A double-strand break (DSBs) occurs in one of the paired DNAs followed by enzymatic trimming back of nucleotides on the new single-strand ends.

A free 3' end invades the unbroken helix and displaces a loop of single strand DNA. A DNA polymerase elongates the free 3' end of the invading strand, further displacing the looped out strand, which then pairs with an exposed single-strand on the opposing helix.

The displaced strand serves as a template for enzymatic extension from the 3' end of the paired single strand, which eventually crosses the junction and switches templates. As 3' and 5' ends meet, strands join to form two Holliday junctions.

There are two ways to resolve each Holliday junction by single cleavage and rejoining, so there are four ways to resolve the double Holliday structure by two cleavages and rejoinings.

Two of these combinations of cleavage and rejoining generate non-crossover recombinants. The other two combinations of cleavage and rejoining generate crossover recombinants. All recombinants include heteroduplex segments.

Double-strand breaks, in which both strands in the double helix are severed, are particularly hazardous to the cell because they can lead to genome rearrangements. Three mechanisms exist to repair DSBs: non-homologous end joining (NHEJ), microhomology-mediated end joining (MMEJ) and homologous recombination.^[14]

In NHEJ, DNA Ligase IV, a specialized DNA ligase that forms a complex with the cofactor XRCC4, directly joins the two ends.^[16] To guide accurate repair, NHEJ relies on short homologous sequences called microhomologies present on the single-stranded tails of the DNA ends to be joined. If these overhangs are compatible, repair is usually accurate.^{[17] [18] [19] [20]} NHEJ can also introduce mutations during repair. Loss of damaged nucleotides at the break site can lead to deletions, and joining of nonmatching termini forms translocations. NHEJ is especially important before the cell has replicated its DNA, since there is no template available for repair by homologous recombination. There are "backup" NHEJ pathways in higher eukaryotes.^[21] Besides its role as a genome caretaker, NHEJ is required for joining hairpin-capped double-strand breaks induced during V(D)J recombination, the process that generates diversity in B-cell and T-cell receptors in the vertebrate immune system.^[22]

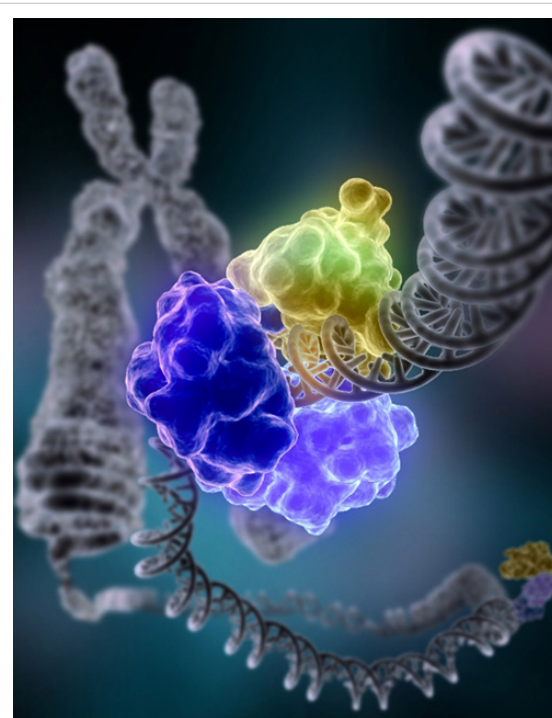
Homologous recombination requires the presence of an identical or nearly identical sequence to be used as a template for repair of the break. The enzymatic machinery responsible for this repair process is nearly identical to the machinery responsible for chromosomal crossover during meiosis. This pathway allows a damaged chromosome to be repaired using a sister chromatid (available in G2 after DNA replication) or a homologous chromosome as a template. DSBs caused by the replication machinery attempting to synthesize across a single-strand break or unrepaired lesion cause collapse of the replication fork and are typically repaired by recombination.

Topoisomerases introduce both single- and double-strand breaks in the course of changing the DNA's state of supercoiling, which is especially common in regions near an open replication fork. Such breaks are not considered DNA damage because they are a natural intermediate in the topoisomerase biochemical mechanism and are immediately repaired by the enzymes that created them.

A team of French researchers bombarded *Deinococcus radiodurans* to study the mechanism of double-strand break DNA repair in that organism. At least two copies of the genome, with random DNA breaks, can form DNA fragments through annealing. Partially overlapping fragments are then used for synthesis of homologous regions through a moving D-loop that can continue extension until they find complementary partner strands. In the final step there is crossover by means of RecA-dependent homologous recombination.^[23]

Translesion synthesis

Translesion synthesis is a DNA damage tolerance process that allows the DNA replication machinery to replicate past DNA lesions such as thymine dimers or AP sites.^[24] It involves switching out regular DNA polymerases for specialized translesion polymerases (e.g. DNA polymerase V), often with larger active sites that can facilitate the insertion of bases opposite damaged nucleotides. The polymerase switching is thought to be mediated by, among other factors, the post-translational modification of the replication processivity factor PCNA. Translesion synthesis polymerases often have low fidelity (high propensity to insert wrong bases) relative to regular polymerases. However, many are extremely efficient at inserting correct bases opposite specific types of damage. For example, Pol η mediates error-free bypass of lesions induced by UV irradiation, whereas Pol ζ introduces mutations at these sites. From a cellular perspective, risking the introduction of point mutations during translesion synthesis may be



DNA ligase, shown above repairing chromosomal damage, is an enzyme that joins broken nucleotides together by catalyzing the formation of an internucleotide ester bond between the phosphate backbone and the deoxyribose nucleotides.

preferable to resorting to more drastic mechanisms of DNA repair, which may cause gross chromosomal aberrations or cell death.

Global response to DNA damage

Cells exposed to ionizing radiation, ultraviolet light or chemicals are prone to acquire multiple sites of bulky DNA lesions and double strand breaks. Moreover, DNA damaging agents can damage other biomolecules such as proteins, carbohydrates, lipids and RNA. The accumulation of damage, specifically double strand breaks or adducts stalling the replication forks, are among known stimulation signals for a global response to DNA damage.^[25] The global response to damage is an act directed toward the cells' own preservation and triggers multiple pathways of macromolecular repair, lesion bypass, tolerance or apoptosis. The common features of global response are induction of multiple genes, cell cycle arrest, and inhibition of cell division.

DNA damage checkpoints

After DNA damage, cell cycle checkpoints are activated. Checkpoint activation pauses the cell cycle and gives the cell time to repair the damage before continuing to divide. DNA damage checkpoints occur at the G1/S and G2/M boundaries. An intra-S checkpoint also exists. Checkpoint activation is controlled by two master kinases, ATM and ATR. ATM responds to DNA double-strand breaks and disruptions in chromatin structure,^[26] whereas ATR primarily responds to stalled replication forks. These kinases phosphorylate downstream targets in a signal transduction cascade, eventually leading to cell cycle arrest. A class of checkpoint mediator proteins including BRCA1, MDC1, and 53BP1 has also been identified.^[27] These proteins seem to be required for transmitting the checkpoint activation signal to downstream proteins.

p53 is an important downstream target of ATM and ATR, as it is required for inducing apoptosis following DNA damage.^[28] At the G1/S checkpoint, p53 functions by deactivating the CDK2/cyclin E complex. Similarly, p21 mediates the G2/M checkpoint by deactivating the CDK1/cyclin B complex.

The prokaryotic SOS response

The SOS response is the term used to describe changes in gene expression in *Escherichia coli* and other bacteria in response to extensive DNA damage. The prokaryotic SOS system is regulated by two key proteins: LexA and RecA. The LexA homodimer is a transcriptional repressor that binds to operator sequences commonly referred to as SOS boxes. In *Escherichia coli* it is known that LexA regulates transcription of approximately 48 genes including the *lexA* and *recA* genes.^[29] The SOS response is known to be widespread in the Bacteria domain, but it is mostly absent in some bacterial phyla, like the Spirochetes.^[30] The most common cellular signals activating the SOS response are regions of single stranded DNA (ssDNA), arising from stalled replication forks or double strand breaks, which are processed by DNA helicase to separate the two DNA strands.^[25] In the initiation step, RecA protein binds to ssDNA in an ATP hydrolysis driven reaction creating RecA–ssDNA filaments. RecA–ssDNA filaments activate LexA autoprotease activity which ultimately leads to cleavage of LexA dimer and subsequent LexA degradation. The loss of LexA repressor induces transcription of the SOS genes and allows for further signal induction, inhibition of cell division and an increase in levels of proteins responsible for damage processing.

In *Escherichia coli*, SOS boxes are 20-nucleotide long sequences near promoters with palindromic structure and a high degree of sequence conservation. In other classes and phyla, the sequence of SOS boxes varies considerably, with different length and composition, but it is always highly conserved and one of the strongest short signals in the genome.^[30] The high information content of SOS boxes permits differential binding of LexA to different promoters and allows for timing of the SOS response. Logically, the lesion repair genes are induced at the beginning of SOS response. The error prone translesion polymerases, for example: UmuCD'2 (also called DNA polymerase V), are induced later on as a last resort.^[31] Once the DNA damage is repaired or bypassed using polymerases or through recombination, the amount of single-stranded DNA in cells is decreased, lowering the amounts of RecA filaments

decreases cleavage activity of LexA homodimer which subsequently binds to the SOS boxes near promoters and restores normal gene expression.

Eukaryotic transcriptional responses to DNA damage

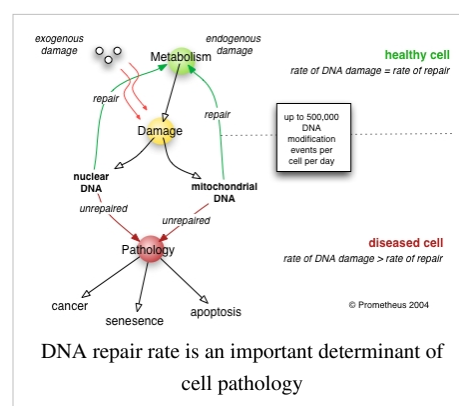
Eukaryotic cells exposed to DNA damaging agents also activate important defensive pathways by inducing multiple proteins involved in DNA repair, cell cycle checkpoint control, protein trafficking and degradation. Such genome wide transcriptional response is very complex and tightly regulated, thus allowing coordinated global response to damage. Exposure of yeast *Saccharomyces cerevisiae* to DNA damaging agents results in overlapping but distinct transcriptional profiles. Similarities to environmental shock response indicates that a general global stress response pathway exist at the level of transcriptional activation. In contrast, different human cell types respond to damage differently indicating an absence of a common global response. The probable explanation for this difference between yeast and human cells may be in the heterogeneity of mammalian cells. In an animal different types of cells are distributed amongst different organs which have evolved different sensitivities to DNA damage.^[31]

In general global response to DNA damage involves expression of multiple genes responsible for postreplication repair, homologous recombination, nucleotide excision repair, DNA damage checkpoint, global transcriptional activation, genes controlling mRNA decay and many others. A large amount of damage to a cell leaves it with an important decision: undergo apoptosis and die, or survive at the cost of living with a modified genome. An increase in tolerance to damage can lead to an increased rate of survival which will allow a greater accumulation of mutations. Yeast Rev1 and human polymerase η are members of [Y family translesion DNA polymerases present during global response to DNA damage and are responsible for enhanced mutagenesis during a global response to DNA damage in eukaryotes.^[25]

DNA repair and aging

Pathological effects of poor DNA repair

Experimental animals with genetic deficiencies in DNA repair often show decreased life span and increased cancer incidence.^[12] For example, mice deficient in the dominant NHEJ pathway and in telomere maintenance mechanisms get lymphoma and infections more often, and consequently have shorter life spans than wild-type mice.^[32] Similarly, mice deficient in a key repair and transcription protein that unwinds DNA helices have premature onset of aging-related diseases and consequent shortening of life span.^[33] However, not every DNA repair deficiency creates exactly the predicted effects; mice deficient in the NER pathway exhibited shortened life span without correspondingly higher rates of mutation.^[34]



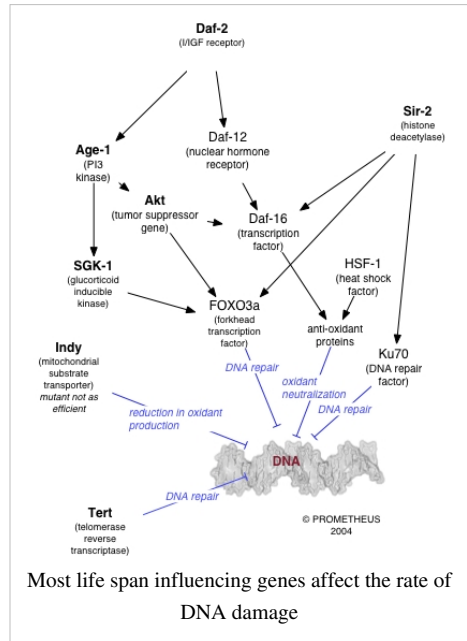
If the rate of DNA damage exceeds the capacity of the cell to repair it, the accumulation of errors can overwhelm the cell and result in early senescence, apoptosis or cancer. Inherited diseases associated with faulty DNA repair functioning result in premature aging,^[12] increased sensitivity to carcinogens, and correspondingly increased cancer risk (see below). On the other hand, organisms with enhanced DNA repair systems, such as *Deinococcus radiodurans*, the most radiation-resistant known organism, exhibit remarkable resistance to the double strand break-inducing effects of radioactivity, likely due to enhanced efficiency of DNA repair and especially NHEJ.^[35]

Longevity and caloric restriction

A number of individual genes have been identified as influencing variations in life span within a population of organisms. The effects of these genes is strongly dependent on the environment, particularly on the organism's diet. Caloric restriction reproducibly results in extended life span in a variety of organisms, likely via nutrient sensing pathways and decreased metabolic rate. The molecular mechanisms by which such restriction results in lengthened life span are as yet unclear (see^[36] for some discussion); however, the behavior of many genes known to be involved in DNA repair is altered under conditions of caloric restriction.

For example, increasing the gene dosage of the gene SIR-2, which regulates DNA packaging in the nematode worm *Caenorhabditis elegans*, can significantly extend life span.^[37] The mammalian homolog of SIR-2 is known to induce downstream DNA repair factors involved in NHEJ, an activity that is especially promoted under conditions of caloric restriction.^[38] Caloric restriction has been closely linked to the rate of base excision repair in the nuclear DNA of rodents,^[39] although similar effects have not been observed in mitochondrial DNA.^[40]

Interestingly, the *C. elegans* gene AGE-1, an upstream effector of DNA repair pathways, confers dramatically extended life span under free-feeding conditions but leads to a decrease in reproductive fitness under conditions of caloric restriction.^[41] This observation supports the pleiotropy theory of the biological origins of aging, which suggests that genes conferring a large survival advantage early in life will be selected for even if they carry a corresponding disadvantage late in life.



Medicine and DNA repair modulation

Hereditary DNA repair disorders

Defects in the NER mechanism are responsible for several genetic disorders, including:

- xeroderma pigmentosum: hypersensitivity to sunlight/UV, resulting in increased skin cancer incidence and premature aging
- Cockayne syndrome: hypersensitivity to UV and chemical agents
- trichothiodystrophy: sensitive skin, brittle hair and nails

Mental retardation often accompanies the latter two disorders, suggesting increased vulnerability of developmental neurons.

Other DNA repair disorders include:

- Werner's syndrome: premature aging and retarded growth
- Bloom's syndrome: sunlight hypersensitivity, high incidence of malignancies (especially leukemias).
- ataxia telangiectasia: sensitivity to ionizing radiation and some chemical agents

All of the above diseases are often called "segmental progerias" ("accelerated aging diseases") because their victims appear elderly and suffer from aging-related diseases at an abnormally young age, while not manifesting all the symptoms of old age.

Other diseases associated with reduced DNA repair function include Fanconi's anemia, hereditary breast cancer and hereditary colon cancer.

DNA repair and cancer

Inherited mutations that affect DNA repair genes are strongly associated with high cancer risks in humans. Hereditary nonpolyposis colorectal cancer (HNPCC) is strongly associated with specific mutations in the DNA mismatch repair pathway. BRCA1 and BRCA2, two famous mutations conferring a hugely increased risk of breast cancer on carriers, are both associated with a large number of DNA repair pathways, especially NHEJ and homologous recombination.

Cancer therapy procedures such as chemotherapy and radiotherapy work by overwhelming the capacity of the cell to repair DNA damage, resulting in cell death. Cells that are most rapidly dividing - most typically cancer cells - are preferentially affected. The side effect is that other non-cancerous but rapidly dividing cells such as stem cells in the bone marrow are also affected. Modern cancer treatments attempt to localize the DNA damage to cells and tissues only associated with cancer, either by physical means (concentrating the therapeutic agent in the region of the tumor) or by biochemical means (exploiting a feature unique to cancer cells in the body).

DNA repair and evolution

The basic processes of DNA repair are highly conserved among both prokaryotes and eukaryotes and even among bacteriophage (viruses that infect bacteria); however, more complex organisms with more complex genomes have correspondingly more complex repair mechanisms.^[42] The ability of a large number of protein structural motifs to catalyze relevant chemical reactions has played a significant role in the elaboration of repair mechanisms during evolution. For an extremely detailed review of hypotheses relating to the evolution of DNA repair, see.^[43]

The fossil record indicates that single celled life began to proliferate on the planet at some point during the Precambrian period, although exactly when recognizably modern life first emerged is unclear. Nucleic acids became the sole and universal means of encoding genetic information, requiring DNA repair mechanisms that in their basic form have been inherited by all extant life forms from their common ancestor. The emergence of Earth's oxygen-rich atmosphere (known as the "oxygen catastrophe") due to photosynthetic organisms, as well as the presence of potentially damaging free radicals in the cell due to oxidative phosphorylation, necessitated the evolution of DNA repair mechanisms that act specifically to counter the types of damage induced by oxidative stress.

Rate of evolutionary change

On some occasions, DNA damage is not repaired, or is repaired by an error-prone mechanism which results in a change from the original sequence. When this occurs, mutations may propagate into the genomes of the cell's progeny. Should such an event occur in a germ line cell that will eventually produce a gamete, the mutation has the potential to be passed on to the organism's offspring. The rate of evolution in a particular species (or, more narrowly, in a particular gene) is a function of the rate of mutation. Consequently, the rate and accuracy of DNA repair mechanisms have an influence over the process of evolutionary change.^[44]

See also

- Direct DNA damage
 - Indirect DNA damage
 - DNA damage theory of aging
 - Accelerated aging disease
 - Aging DNA
 - Cell cycle
 - DNA replication
 - Gene therapy
 - Life extension
-

- Human mitochondrial genetics
- Progeria
- Senescence
- The scientific journal *DNA Repair* under Mutation Research

External links

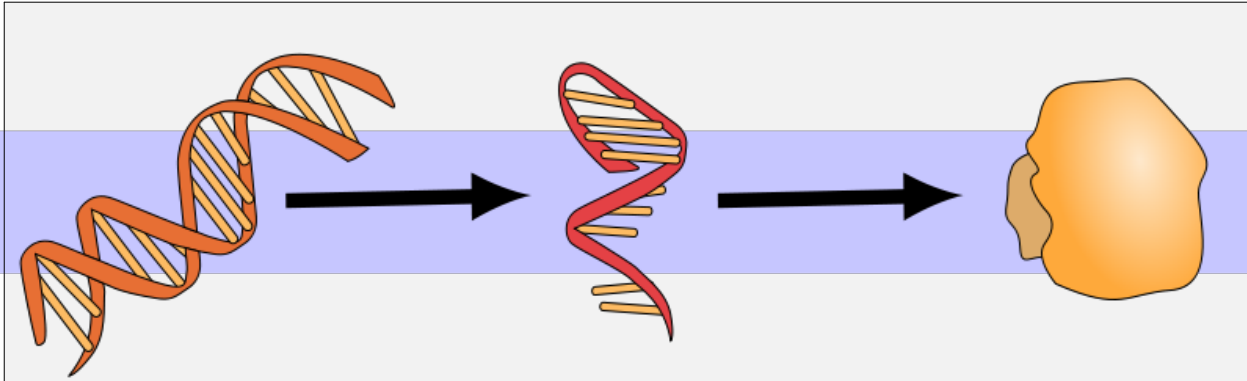
- Roswell Park Cancer Institute DNA Repair Lectures ^[45]
- DNA Repair - A summary of the primary mechanisms ^[46]
- A comprehensive list of Human DNA Repair Genes ^[47]
- 3D structures of some DNA repair enzymes ^[48]
- Human DNA repair diseases ^[49]
- DNA repair special interest group ^[50]
- DNA Repair ^[51]
- DNA Damage and DNA Repair ^[52]
- Segmental Progeria ^[53]
- DNA-damage repair; the good, the bad, and the ugly ^[54]

References

- [1] Lodish H, Berk A, Matsudaira P, Kaiser CA, Krieger M, Scott MP, Zipursky SL, Darnell J. (2004). *Molecular Biology of the Cell*, p963. WH Freeman: New York, NY. 5th ed.
- [2] Acharya PV; The isolation and partial characterization of age-correlated oligo-deoxyribo-ribonucleotides with covalently linked aspartyl-glutamyl polypeptides.(June, 1971). *Johns Hopkins Med J Suppl*, p254-260. PMID 5055816.
- [3] Bjorksten, J; Acharya, PV; Ashman, S; Wetlaufer, DB. Gerogenic Fractions in the Tritiated Rat (July, 1971). *Journal of the American Geriatrics Society*, p561-574; PMID 5106728.
- [4] Browner WS, Kahn AJ, Ziv E, Reiner AP, Oshima J, Cawthon RM, Hsueh WC, Cummings SR. (2004). The genetics of human longevity. *Am J Med* 117(11):851–60.
- [5] Roulston A, Marcellus RC, Branton PE (1999). "Viruses and apoptosis" (http://arjournals.annualreviews.org/doi/abs/10.1146/annurev.micro.53.1.577?url_ver=Z39.88-2003&rft_id=ori:rid:crossref.org&rft_dat=cr_pub=ncbi.nlm.nih.gov). *Annu. Rev. Microbiol.* **53**: 577–628. doi:10.1146/annurev.micro.53.1.577. PMID 10547702. . Retrieved 2008-12-20.
- [6] Madigan MT, Martino JM (2006). *Brock Biology of Microorganisms* (11th ed.). Pearson. pp. 136. ISBN 0-13-196893-9.
- [7] Toshihiro Ohta, Shin-ichi Tokishita, Kayo Mochizuki, Jun Kawase, Masahide Sakahira and Hideo Yamagata, UV Sensitivity and Mutagenesis of the Extremely Thermophilic Eubacterium *Thermus thermophilus* HB27, *Genes and Environment* Vol. 28 (2006), No. 2 p.56–61.
- [8] DNA Lesions That Require Repair: <http://www.ncbi.nlm.nih.gov/books/bv.fcgi?highlight=lesion&rid=mcb.table.3236>
- [9] Braig M, Schmitt CA. (2006) . Oncogene-induced senescence: putting the brakes on tumor development. *Cancer Res* 66: 2881–2884.
- [10] Lynch MD. (2006). How does cellular senescence prevent cancer? *DNA Cell Biol* 25(2):69–78.
- [11] Campisi J, d'Adda di Fagagna F (2007). "Cellular senescence: when bad things happen to good cells.". *Rev Mol Cell Biol.* **8** (9): 729–40. doi:10.1038/nrm2233. PMID 17667954.
- [12] Best,BP (2009). "Nuclear DNA damage as a direct cause of aging" (http://www.benbest.com/lifeext/Nuclear_DNA_in_Aging.pdf) (PDF). *Rejuvenation Research* **12** (3): 199–208. doi:10.1089/rej.2009.0847. PMID 19594328. PMID 19594328. .
- [13] Sancar A. (2003). Structure and function of DNA photolyase and cryptochrome blue-light photoreceptors. *Chem Rev* 103(6):2203–37. PMID 12797829
- [14] Watson JD, Baker TA, Bell SP, Gann A, Levine M, Losick R. (2004). *Molecular Biology of the Gene*, ch. 9 and 10. Peason Benjamin Cummings; CSHL Press. 5th ed.
- [15] Volkert MR. (1988). Adaptive response of *Escherichia coli* to alkylation damage*Environ Mol Mutagen* 11(2):241-55.
- [16] Wilson, T. E., Grawunder, U., and Lieber, M. R. Yeast DNA ligase IV mediates non-homologous DNA end joining. (1997) *Nature* 388, 495–498. PMID 9242411
- [17] Moore JK, Haber JE. Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in *Saccharomyces cerevisiae*. *Mol Cell Biol.* 1996 May;16(5):2164–73. PMID 8628283
- [18] Boulton SJ, Jackson SP. *Saccharomyces cerevisiae* Ku70 potentiates illegitimate DNA double-strand break repair and serves as a barrier to error-prone DNA repair pathways. *EMBO J.* 1996 Sep 16;15(18):5093-103. PMID 8890183
- [19] Wilson, T. E., and Lieber, M. R. Efficient processing of DNA ends during yeast nonhomologous end joining. Evidence for a DNA polymerase beta (Pol4)-dependent pathway. (1999) *J. Biol. Chem.* 274, 23599–23609. PMID 10438542

- [20] Budman J, Chu G. Processing of DNA for nonhomologous end-joining by cell-free extract. *EMBO J.* 2005 Feb 23;24(4):849-60. PMID: 15692565
- [21] Wang H, Perrault AR, Takeda Y, Qin W, Wang H, Iliakis G. (2003). Biochemical evidence for Ku-independent backup pathways of NHEJ. *Nucleic Acids Res* 31(18):5377-88.
- [22] Jung D, Alt FW. Unraveling V(D)J recombination; insights into gene regulation. *Cell.* 2004 Jan 23;116(2):299-311. Review. PMID: 14744439
- [23] Zahradka K, Slade D, Bailone A, Sommer S, Auerbeck D, Petranovic M, Lindner AB, Radman M (2006). "Reassembly of shattered chromosomes in *Deinococcus radiodurans*". *NATURE* **443** (7111): 569-573. doi:10.1038/nature05160. PMID 17006450.
- [24] Waters LS, Minesinger BK, Wiltout ME, D'Souza S, Woodruff RV, Walker GC (March 2009). "Eukaryotic translesion polymerases and their roles and regulation in DNA damage tolerance" (<http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=2650891>). *Microbiol. Mol. Biol. Rev.* **73** (1): 134-54. doi:10.1128/MMBR.00034-08. PMID 19258535. PMC 2650891.
- [25] Friedberg EC, Walker GC, Siede W, Wood RD, Schultz RA, Ellenberger T. (2006). DNA Repair and Mutagenesis, part 3. ASM Press. 2nd ed.
- [26] Bakkenist CJ, Kastan MB. DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation. *Nature.* 2003 Jan 30;421(6922):499-506.
- [27] Wei, Qingyi; Lei Li, David Chen (2007). *DNA Repair, Genetic Instability, and Cancer*. World Scientific. ISBN 9812700145.
- [28] Schonthal, Axel H. (2004). *Checkpoint Controls and Cancer*. Humana Press. ISBN 1588295001.
- [29] Janion C. (2001). Some aspects of the SOS response system-a critical survey. *Acta Biochim Pol.* 48(3):599-610
- [30] Erill, I. Campoy, S. Barbé, J. (2007). Aeons of distress: an evolutionary perspective on the bacterial SOS response. *FEMS Microbiol Rev.* 31(6):637-656.
- [31] Schlacher K, Pham P, Cox MM, and Goodman MF. (2006). Roles of DNA Polymerase V and RecA Protein in SOS Damage-Induced Mutation. *Chem. Rev* 106(2) pp 406-419
- [32] Espejel S, Martin M, Klatt P, Martin-Caballero J, Flores JM, Blasco MA. (2004). Shorter telomeres, accelerated ageing and increased lymphoma in DNA-PKcs-deficient mice. *EMBO Rep* 5(5):503-9.
- [33] de Boer J, Andressoo JO, de Wit J, Huijman J, Beems RB, van Steeg H, Weeda G, van der Horst GT, van Leeuwen W, Themmen AP, Meradji M, Hoeijmakers JH. (2002). Premature aging in mice deficient in DNA repair and transcription. *Science* 296(5571):1276-9.
- [34] Dolle ME, Busuttill RA, Garcia AM, Wijnhoven S, van Drunen E, Niedernhofer LJ, van der Horst G, Hoeijmakers JH, van Steeg H, Vijg J. (2006). Increased genomic instability is not a prerequisite for shortened life span in DNA repair deficient mice. *Mutation Research* 596(1-2):22-35.
- [35] Kobayashi Y, Narumi I, Satoh K, Funayama T, Kikuchi M, Kitayama S, Watanabe H. (2004). Radiation response mechanisms of the extremely radioresistant bacterium *Deinococcus radiodurans*. *Biol Sci Space* 18(3):134-5.
- [36] Spindler SR. (2005). Rapid and reversible induction of the longevity, anticancer and genomic effects of caloric restriction. *Mech Ageing Dev* 126(9):960-6.
- [37] Tissenbaum HA, Guarente L. (2001). Increased dosage of a sir-2 gene extends life span in *Caenorhabditis elegans*. *Nature* 410(6825):227-30.
- [38] Cohen HY, Miller C, Bitterman KJ, Wall NR, Hekking B, Kessler B, Howitz KT, Gorospe M, de Cabo R, Sinclair DA. (2004). Calorie restriction promotes mammalian cell survival by inducing the SIRT1 deacetylase. *Science* 305(5682):390-2.
- [39] Cabellof DC, Yanamadala S, Raffoul JJ, Guo Z, Soofi A, Heydari AR. (2003). Caloric restriction promotes genomic stability by induction of base excision repair and reversal of its age-related decline. *DNA Repair (Amst.)* 2(3):295-307.
- [40] Stuart JA, Karahalil B, Hogue BA, Souza-Pinto NC, Bohr VA. (2004). Mitochondrial and nuclear DNA base excision repair are affected differently by caloric restriction. *FASEB J* 18(3):595-7.
- [41] Walker DW, McColl G, Jenkins NL, Harris J, Lithgow GJ. (2000). Evolution of lifespan in *C. elegans*. *Nature* 405(6784):296-7.
- [42] Cromie GA, Connelly JC, Leach DR (2001). Recombination at double-strand breaks and DNA ends: conserved mechanisms from phage to humans. *Mol Cell.* 8(6):1163-74.
- [43] O'Brien PJ. (2006). Catalytic promiscuity and the divergent evolution of DNA repair enzymes. *Chem Rev* 106(2):720-52.
- [44] Maresca B, Schwartz JH (2006). Sudden origins: a general mechanism of evolution based on stress protein concentration and rapid environmental change. *Anat Rec B New Anat.* Jan;289(1):38-46
- [45] http://asajj.roswellpark.org/huberman/DNA_Repair/DNA_Repair.htm
- [46] <http://whyweage.com/node/20>
- [47] http://www.cgal.icnet.uk/DNA_Repair_Genes.html
- [48] <http://www.biochem.ucl.ac.uk/bsm/xtal/teach/repair/tibs3.html>
- [49] http://www.scielo.br/scielo.php?pid=S0100-84551997000400032&script=sci_arttext&tlng=en
- [50] http://tango01.cit.nih.gov/sig/home.taf?_function=main&SIGInfo_SIGID=32
- [51] <http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/D/DNArepair.html>
- [52] <http://www.benbest.com/lifeext/aging.html#dna>
- [53] <http://www.benbest.com/lifeext/aging.html#progeria>
- [54] https://www.researchgate.net/publication/5565866_DNA-damage_repair_the_good_the_bad_and_the_ugly

DNA Translation



This

article is part of the series on:
Gene expression
a Molecular biology topic (*portal*)
(Glossary)

Introduction to Genetics
General flow: DNA > RNA > Protein
special transfers (RNA > RNA, RNA > DNA, Protein > Protein)
Genetic code
Transcription
Transcription (<i>Transcription factors, RNA Polymerase, promoter</i>) Prokaryotic / Archaeal / Eukaryotic
post-transcriptional modification (<i>hnRNA, Splicing</i>)
Translation
Translation (<i>Ribosome, tRNA</i>) Prokaryotic / Archaeal / Eukaryotic
post-translational modification (<i>functional groups, peptides, structural changes</i>)
gene regulation
epigenetic regulation (<i>Genomic imprinting</i>)
transcriptional regulation
post-transcriptional regulation (<i>sequestration, alternative splicing, miRNA</i>)
translational regulation
post-translational regulation (<i>reversible, irreversible</i>)
ask a question ^[1] , edit ^[2]

Translation is the first stage of protein biosynthesis (part of the overall process of gene expression). In translation, messenger RNA (mRNA) produced in transcription is decoded to produce a specific amino acid chain, or

polypeptide, that will later fold into an active protein. Translation occurs in the cell's cytoplasm, where the large and small subunits of the ribosome are located, and bind to the mRNA. The ribosome facilitates decoding by inducing the binding of tRNAs with complementary anticodon sequences to that of the mRNA. The tRNAs carry specific amino acids that are chained together into a polypeptide as the mRNA passes through and is "read" by the ribosome in a fashion reminiscent to that of a stock ticker and ticker tape.

In many instances, the entire ribosome/mRNA complex will bind to the outer membrane of the rough endoplasmic reticulum and release the nascent protein polypeptide inside for later vesicle transport and secretion outside of the cell. Many types of transcribed RNA, such as transfer RNA, ribosomal RNA, and small nuclear RNA, do not undergo translation into proteins.

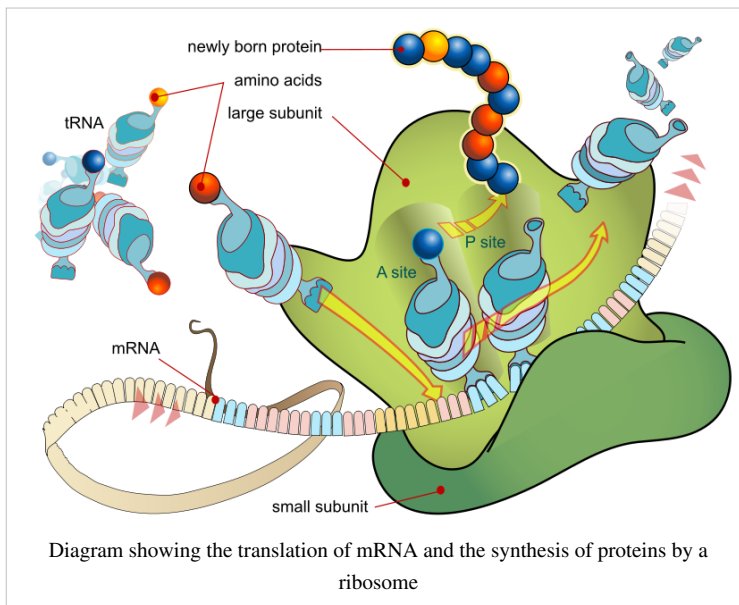
Translation proceeds in four phases: activation, initiation, elongation and termination (all describing the growth of the amino acid chain, or polypeptide that is the product of translation). Amino acids are brought to ribosomes and assembled into proteins.

In activation, the correct amino acid is covalently bonded to the correct transfer RNA (tRNA). The amino acid is joined by its carboxyl group to the 3' OH of the tRNA by a peptide bond. When the tRNA has an amino acid linked to it, it is termed "charged". Initiation involves the small subunit of the ribosome binding to 5' end of mRNA with the help of initiation factors (IF). Termination of the polypeptide happens when the A site of the ribosome faces a stop codon (UAA, UAG, or UGA). No tRNA can recognize or bind to this codon. Instead, the stop codon induces the binding of a release factor protein that prompts the disassembly of the entire ribosome/mRNA complex.

A number of antibiotics act by inhibiting translation; these include anisomycin, cycloheximide, chloramphenicol, tetracycline, streptomycin, erythromycin, and puromycin, among others. Prokaryotic ribosomes have a different structure from that of eukaryotic ribosomes, and thus antibiotics can specifically target bacterial infections without any detriment to a eukaryotic host's cells.

Basic mechanisms

See articles at prokaryotic translation and eukaryotic translation



The mRNA carries genetic information encoded as a ribonucleotide sequence from the chromosomes to the ribosomes. The ribonucleotides are "read" by translational machinery in a sequence of nucleotide triplets called codons. Each of those triplets codes for a specific amino acid.

The ribosome molecules translate this code to a specific sequence of amino acids. The ribosome is a multisubunit structure containing rRNA and proteins. It is the "factory" where amino acids are assembled into proteins. tRNAs are small noncoding RNA chains (74-93 nucleotides) that transport amino acids to the ribosome. tRNAs have a site for amino acid

attachment, and a site called an anticodon. The anticodon is an RNA triplet complementary to the mRNA triplet that codes for their cargo amino acid.

Aminoacyl tRNA synthetase (an enzyme) catalyzes the bonding between specific tRNAs and the amino acids that their anticodons sequences call for. The product of this reaction is an aminoacyl-tRNA molecule. This

12: The Alternative Yeast Nuclear Code
13: The Ascidian Mitochondrial Code
14: The Alternative Flatworm Mitochondrial Code
15: Blepharisma Nuclear Code
16: Chlorophycean Mitochondrial Code
21: Trematode Mitochondrial Code
22: Scenedesmus obliquus mitochondrial Code
23: Thraustochytrium Mitochondrial Code

Software examples

- ApE ^[6] (Mac, Windows, Unix)
- Serial Cloner ^[7] (A DNA editing and manipulating software for MacOS and Windows)
- DNA Strider (Mac)
- ExPASy Translate Tool ^[8] (webserver)
- Virtual Ribosome ^[9] (webserver, cross-platform command-line)
- DNA to protein translation ^[10] (webserver, 13 genomic codes or custom ones)

Example of computational translation - notice the indication of (alternative) start-codons:

```

VIRTUAL RIBOSOME
-----
Translation table: Standard SGC0

>Seq1
Reading frame: 1

      M  V  L  S  A  A  D  K  G  N  V  K  A  A  W  G  K  V  G  G  H  A  A
    E  Y  G  A  E  A  L
5'
ATGGTGCTGTCTGCCGCCGACAAGGGCAATGTCAAGGCCGCCTGGGGCAAGGTTGGCGGCCACGCTGCAGAGTATGGCGCAGAGGCCCT
90

>>>...)).....))

      E  R  M  F  L  S  F  P  T  T  K  T  Y  F  P  H  F  D  L  S  H  G  S
    A  Q  V  K  G  H  G
5'
GAGAGGATGTTCTTGAGCTTCCCCACCACCAAGACCTACTTCCCCCACTTCGACCTGAGCCACGGCTCCGCGCAGGTCAAGGGCCACGG
180

.....>>>...)).....))

      A  K  V  A  A  A  L  T  K  A  V  E  H  L  D  D  L  P  G  A  L  S  E
    L  S  D  L  H  A  H
5'
GCGAAGGTGGCCGCCGCGCTGACCAAAGCGGTGGAACACCTGGACGACCTGCCCGGTGCCCTGTCTGAACTGAGTGACCTGCACGCTCA

```

270

.....))).....))).....))).....))).....))).....))).....

K L R V D P V N F K L L S H S L L V T L A S H
L P S D F T P

5'

AAGCTGCGTGTGGACCCGGTCAACTTCAAGCTTCTGAGCCACTCCCTGCTGGTGACCCTGGCCTCCACCTCCCCAGTGATTTCACCCO

360

...))).....))).....))).....))).....))).....

A V H A S L D K F L A N V S T V L T S K Y R *

5'

GCGGTCCACGCCTCCCTGGACAAGTTCTTGGCCAACGTGAGCACCGTGCTGACCTCCAAATACCGTTAA

429

.....))).....))).....))).....**

Annotation key:

>>> : START codon (strict)

))) : START codon (alternative)

*** : STOP

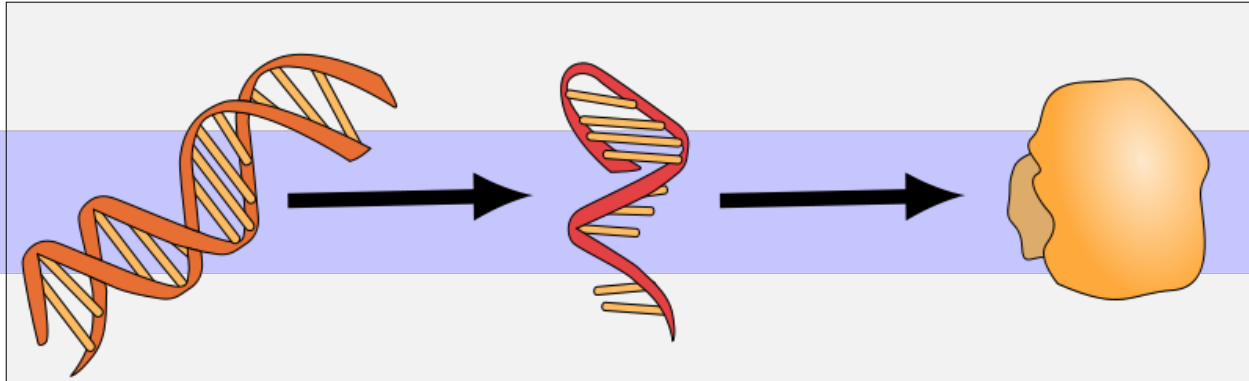
See also

- Expanded genetic code
- Protein methods

References

- http://en.wikipedia.org/w/index.php?title=Wikipedia_talk:WikiProject_Molecular_and_Cellular_Biology&action=edit§ion=new
 - <http://en.wikipedia.org/w/index.php?title=Template:MolBioGeneExp&action=edit>
 - Ross JF, Orlowski M (February 1982). "Growth-rate-dependent adjustment of ribosome function in chemostat-grown cells of the fungus *Mucor racemosus*" (<http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=216554>). *J. Bacteriol.* **149** (2): 650–3. PMID 6799491. PMC 216554.
 - <http://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi?mode=c>
 - <http://www.ncbi.nlm.nih.gov/Taxonomy/>
 - <http://www.biology.utah.edu/jorgensen/wayned/ape/>
 - http://www.serialbasics.com/Serial_Cloner.html
 - <http://www.expasy.org/tools/dna.html>
 - <http://www.cbs.dtu.dk/services/VirtualRibosome/>
 - <http://insilico.ehu.es/translate/>
- Champe, Pamela C; Harvey, Richard A; Ferrier, Denise R (2004). *Lippincott's Illustrated Reviews: Biochemistry* (3rd ed.). Hagerstwon, MD: Lippincott Williams & Wilkins. ISBN 0-7817-2265-9.
 - Cox, Michael; Nelson, David R.; Lehninger, Albert L (2005). *Lehninger principles of biochemistry* (4th ed.). San Francisco...: W.H. Freeman. ISBN 0-7167-4339-6.

DNA Transcription



This

article is part of the series on:
Gene expression
a Molecular biology topic (*portal*)
(Glossary)

Introduction to Genetics
General flow: DNA > RNA > Protein
special transfers (RNA > RNA, RNA > DNA, Protein > Protein)
Genetic code
Transcription
Transcription (<i>Transcription factors</i> , <i>RNA Polymerase,promoter</i>) Prokaryotic / Archaeal / Eukaryotic
post-transcriptional modification (<i>hnRNA,Splicing</i>)
Translation
Translation (<i>Ribosome,tRNA</i>) Prokaryotic / Archaeal / Eukaryotic
post-translational modification (<i>functional groups, peptides</i> , <i>structural changes</i>)
gene regulation
epigenetic regulation (<i>Genomic imprinting</i>)
transcriptional regulation
post-transcriptional regulation (<i>sequestration</i> , <i>alternative splicing,miRNA</i>)
translational regulation
post-translational regulation (<i>reversible,irreversible</i>)
ask a question ^[1] , edit ^[2]

Transcription, or **RNA synthesis**, is the process of creating an equivalent RNA copy of a sequence of DNA^[1] . Both RNA and DNA are nucleic acids, which use base pairs of nucleotides as a complementary language that can be

converted back and forth from DNA to RNA in the presence of the correct enzymes. During transcription, a DNA sequence is read by RNA polymerase, which produces a complementary, antiparallel RNA strand. As opposed to DNA replication, transcription results in an RNA complement that includes uracil (U) in all instances where thymine (T) would have occurred in a DNA complement.

Transcription is the first step leading to gene expression. The stretch of DNA transcribed into an RNA molecule is called a *transcription unit* and encodes at least one gene. If the gene transcribed encodes for a protein, the result of transcription is messenger RNA (mRNA), which will then be used to create that protein via the process of translation. Alternatively, the transcribed gene may encode for either ribosomal RNA (rRNA) or transfer RNA (tRNA), other components of the protein-assembly process, or other ribozymes.

A DNA transcription unit encoding for a protein contains not only the sequence that will eventually be directly translated into the protein (the *coding sequence*) but also *regulatory sequences* that direct and regulate the synthesis of that protein. The regulatory sequence before (upstream from) the coding sequence is called the five prime untranslated region (5'UTR), and the sequence following (downstream from) the coding sequence is called the three prime untranslated region (3'UTR).

Transcription has some proofreading mechanisms, but they are fewer and less effective than the controls for copying DNA; therefore, transcription has a lower copying fidelity than DNA replication.^[2]

As in DNA replication, DNA is read from 3' → 5' during transcription. Meanwhile, the complementary RNA is created from the 5' → 3' direction. Although DNA is arranged as two antiparallel strands in a double helix, only one of the two DNA strands, called the template strand, is used for transcription. This is because RNA is only single-stranded, as opposed to double-stranded DNA. The other DNA strand is called the coding strand, because its sequence is the same as the newly created RNA transcript (except for the substitution of uracil for thymine). The use of only the 3' → 5' strand eliminates the need for the Okazaki fragments seen in DNA replication.

Transcription is divided into 5 stages: *pre-initiation*, *initiation*, *promoter clearance*, *elongation* and *termination*.

Major steps

Pre-initiation

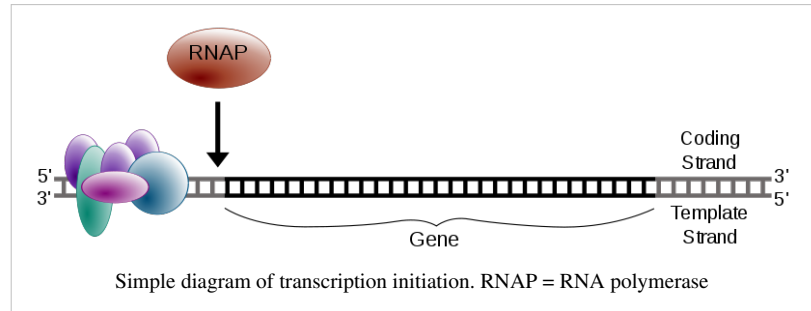
In eukaryotes, RNA polymerase, and therefore the initiation of transcription, requires the presence of a core promoter sequence in the DNA. Promoters are regions of DNA which promote transcription and in eukaryotes, are found at -30, -75 and -90 base pairs upstream from the start site of transcription. Core promoters are sequences within the promoter which are essential for transcription initiation. RNA polymerase is able to bind to core promoters in the presence of various specific transcription factors.

The most common type of core promoter in eukaryotes is a short DNA sequence known as a TATA box, found -30 base pairs from the start site of transcription. The TATA box, as a core promoter, is the binding site for a transcription factor known as TATA binding protein (TBP), which is itself a subunit of another transcription factor, called Transcription Factor II D (TFIID). After TFIID binds to the TATA box via the TBP, five more transcription factors and RNA polymerase combine around the TATA box in a series of stages to form a preinitiation complex. One transcription factor, DNA helicase, has helicase activity and so is involved in the separating of opposing strands of double-stranded DNA to provide access to a single-stranded DNA template. However, only a low, or basal, rate of transcription is driven by the preinitiation complex alone. Other proteins known as activators and repressors, along with any associated coactivators or corepressors, are responsible for modulating transcription rate.

The transcription preinitiation in archaea, formerly a domain of prokaryote, is essentially homologous to that of eukaryotes, but is much less complex.^[3] The archaeal preinitiation complex assembles at a TATA-box binding site; however, in archaea, this complex is composed of only RNA polymerase II, TBP, and TFB (the archaeal homologue of eukaryotic transcription factor II B (TFIIB)).^{[4] [5]}

Initiation

In bacteria, a domain of prokaryotes, transcription begins with the binding of RNA polymerase to the promoter in DNA. RNA polymerase is a core enzyme consisting of five subunits: 2 α subunits, 1 β subunit, 1 β' subunit, and 1 ω subunit. At the start of initiation, the core enzyme is associated with a sigma factor (number 70) that aids in finding the appropriate -35 and -10 base pairs downstream of promoter sequences.



Transcription initiation is more complex in eukaryotes. Eukaryotic RNA polymerase does not directly recognize the core promoter sequences. Instead, a collection of proteins called transcription factors mediate the binding of RNA polymerase and the initiation of transcription. Only after certain transcription factors are attached to the promoter does the RNA polymerase bind to it. The completed assembly of transcription factors and RNA polymerase bind to the promoter, forming a transcription initiation complex. Transcription in the archaea domain is similar to transcription in eukaryotes.^[6]

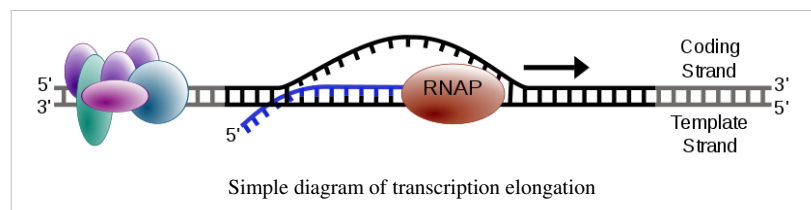
Promoter clearance

After the first bond is synthesized, the RNA polymerase must clear the promoter. During this time there is a tendency to release the RNA transcript and produce truncated transcripts. This is called *abortive initiation* and is common for both eukaryotes and prokaryotes^[7]. Abortive initiation continues to occur until the σ factor rearranges, resulting in the transcription elongation complex (which gives a 35 bp moving footprint). The σ factor is released before 80 nucleotides of mRNA are synthesized^[8]. Once the transcript reaches approximately 23 nucleotides, it no longer slips and elongation can occur. This, like most of the remainder of transcription, is an energy-dependent process, consuming adenosine triphosphate (ATP).

Promoter clearance coincides with phosphorylation of serine 5 on the carboxy terminal domain of RNA Pol in eukaryotes, which is phosphorylated by TFIIH.

Elongation

One strand of DNA, the *template strand* (or noncoding strand), is used as a template for RNA synthesis. As transcription proceeds, RNA polymerase traverses the template strand and uses base pairing



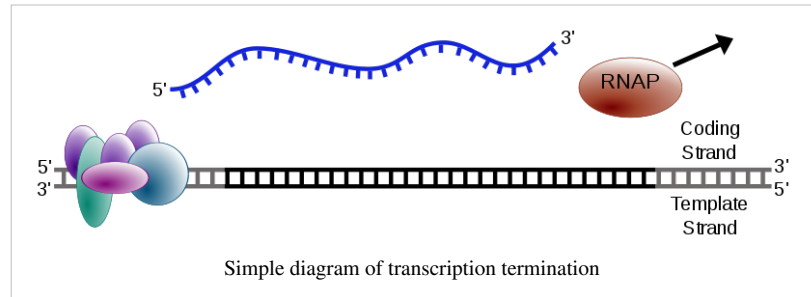
complementarity with the DNA template to create an RNA copy. Although RNA polymerase traverses the template strand from 3' $\rightarrow$ 5', the coding (non-template) strand and newly-formed RNA can also be used as reference points, so transcription can be described as occurring 5' $\rightarrow$ 3'. This produces an RNA molecule from 5' $\rightarrow$ 3', an exact copy of the coding strand (except that thymines are replaced with uracils, and the nucleotides are composed of a ribose (5-carbon) sugar where DNA has deoxyribose (one less oxygen atom) in its sugar-phosphate backbone).

Unlike DNA replication, mRNA transcription can involve multiple RNA polymerases on a single DNA template and multiple rounds of transcription (amplification of particular mRNA), so many mRNA molecules can be rapidly produced from a single copy of a gene.

Elongation also involves a proofreading mechanism that can replace incorrectly incorporated bases. In eukaryotes, this may correspond with short pauses during transcription that allow appropriate RNA editing factors to bind. These pauses may be intrinsic to the RNA polymerase or due to chromatin structure.

Termination

Bacteria use two different strategies for transcription termination. In Rho-independent transcription termination, RNA transcription stops when the newly synthesized RNA molecule forms a G-C rich hairpin loop followed by a run of U's, which makes it detach from the DNA template.



In the "Rho-dependent" type of termination, a protein factor called "Rho" destabilizes the interaction between the template and the mRNA, thus releasing the newly synthesized mRNA from the elongation complex.

Transcription termination in eukaryotes is less understood but involves cleavage of the new transcript followed by template-independent addition of As at its new 3' end, in a process called polyadenylation.

Measuring and detecting transcription

Transcription can be measured and detected in a variety of ways:

- Nuclear Run-on assay: measures the relative abundance of newly formed transcripts
- RNase protection assay and ChIP-Chip of RNAP: detect active transcription sites
- RT-PCR: measures the absolute abundance of total or nuclear RNA levels, which may however differ from transcription rates
- DNA microarrays: measures the relative abundance of the global total or nuclear RNA levels; however, these may differ from transcription rates
- In situ hybridization: detects the presence of a transcript
- MS2 tagging: by incorporating RNA stem loops, such as MS2, into a gene, these become incorporated into newly synthesized RNA. The stem loops can then be detected using a fusion of GFP and the MS2 coat protein, which has a high affinity, sequence specific interaction with the MS2 stem loops. The recruitment of GFP to the site of transcription is visualised as a single fluorescent spot. This remarkable new approach has revealed that transcription occurs in discontinuous bursts, or pulses (see Transcriptional bursting). With the notable exception of in situ techniques, most other methods provide cell population averages, and are not capable of detecting this fundamental property of genes^[9].
- Northern blot: the traditional method, and until the advent of RNA-Seq, the most quantitative
- RNA-Seq: applies next-generation sequencing techniques to sequence whole transcriptomes, which allows the measurement of relative abundance of RNA, as well as the detection of additional variations such as fusion genes, post-translational edits and novel splice sites

Transcription factories

Active transcription units are clustered in the nucleus, in discrete sites called transcription factories or euchromatin. Such sites can be visualized by allowing engaged polymerases to extend their transcripts in tagged precursors (Br-UTP or Br-U) and immuno-labeling the tagged nascent RNA. Transcription factories can also be localized using fluorescence in situ hybridization or marked by antibodies directed against polymerases. There are ~10,000 factories in the nucleoplasm of a HeLa cell, among which are ~8,000 polymerase II factories and ~2,000 polymerase III

factories. Each polymerase II factory contains ~8 polymerases. As most active transcription units are associated with only one polymerase, each factory usually contains ~8 different transcription units. These units might be associated through promoters and/or enhancers, with loops forming a 'cloud' around the factor.

History

A molecule which allows the genetic material to be realized as a protein was first hypothesized by François Jacob and Jacques Monod. RNA synthesis by RNA polymerase was established *in vitro* by several laboratories by 1965; however, the RNA synthesized by these enzymes had properties that suggested the existence of an additional factor needed to terminate transcription correctly.

In 1972, Walter Fiers became the first person to actually prove the existence of the terminating enzyme.

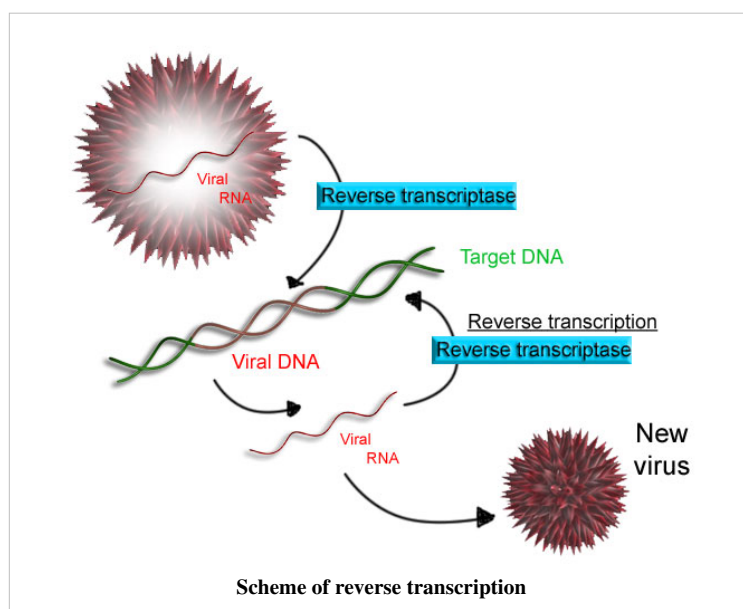
Roger D. Kornberg won the 2006 Nobel Prize in Chemistry "for his studies of the molecular basis of eukaryotic transcription".^[10]

Reverse transcription

Some viruses (such as HIV, the cause of AIDS), have the ability to transcribe RNA into DNA. HIV has an RNA genome that is duplicated into DNA. The resulting DNA can be merged with the DNA genome of the host cell. The main enzyme responsible for synthesis of DNA from an RNA template is called reverse transcriptase. In the case of HIV, reverse transcriptase is responsible for synthesizing a complementary DNA strand (cDNA) to the viral RNA genome. An associated enzyme, ribonuclease H, digests the RNA strand, and reverse transcriptase synthesizes a complementary strand of DNA to form a double helix DNA structure. This cDNA is integrated into the host cell's

genome via another enzyme (integrase) causing the host cell to generate viral proteins which reassemble into new viral particles. Subsequently, the host cell undergoes programmed cell death, apoptosis.

Some eukaryotic cells contain an enzyme with reverse transcription activity called telomerase. Telomerase is a reverse transcriptase that lengthens the ends of linear chromosomes. Telomerase carries an RNA template from which it synthesizes DNA repeating sequence, or "junk" DNA. This repeated sequence of DNA is important because every time a linear chromosome is duplicated it is shortened in length. With "junk" DNA at the ends of chromosomes, the shortening eliminates some of the non-essential, repeated sequence rather than the protein-encoding DNA sequence farther away from the chromosome end. Telomerase is often activated in cancer cells to enable cancer cells to duplicate their genomes indefinitely without losing important protein-coding DNA sequence. Activation of telomerase could be part of the process that allows cancer cells to become technically immortal. However, the true *in vivo* significance of telomerase has still not been empirically proven.



See also

- Genetics
- Molecular biology
- Translation - process of decoding RNA to form polypeptides.
- Splicing - process of removing introns from precursor messenger RNA (pre-mRNA) to make form messenger RNA (mRNA).
- Reverse transcription - process viruses use to make DNA from RNA
- Crick's central dogma - DNA is transcribed to RNA which is translated to polypeptides, never the other way around.

Further reading

- Lehninger Principles of Biochemistry, 5th edition, David L. Nelson & Michael M. Cox
- Principles of Nuclear Structure and Function, Peter R. Cook
- Essential Genetics, Peter J. Russell

External links

- Interactive Java simulation of transcription initiation.^[11] From Center for Models of Life^[12] at the Niels Bohr Institute.
- Interactive Java simulation of transcription interference--a game of promoter dominance in bacterial virus.^[13] From Center for Models of Life^[12] at the Niels Bohr Institute.
- Biology animations about this topic under Chapter 15 and Chapter 18^[14]

References

- [1] MedicineNet.com. "Transcription definition" (<http://www.medterms.com/script/main/art.asp?articlekey=5835>). . Retrieved 11 October 2009.
- [2] Berg J, Tymoczko JL, Stryer L (2006). *Biochemistry* (6th ed.). San Francisco: W. H. Freeman. ISBN 0716787245.
- [3] Littlefield, O., Korkhin, Y., and Sigler, P.B. (1999). "The structural basis for the oriented assembly of a TBP/TFB/promoter complex". *PNAS* **96**: 13668–13673. doi:10.1073/pnas.96.24.13668.
- [4] Hausner, W; Thomm, M (2001). "Events during Initiation of Archaeal Transcription: Open Complex Formation and DNA-Protein Interactions". *Journal of Bacteriology* **183** (10): 3025–3031. doi:10.1128/JB.183.10.3025-3031.2001. PMID 11325929.
- [5] Qureshi, SA; Bell, SD; Jackson, SP (1997). "Factor requirements for transcription in the archaeon *Sulfolobus shibatae*". *EMBO Journal* **16** (10): 2927–2936. doi:10.1093/emboj/16.10.2927. PMID 9184236.
- [6] Mohamed Ouhammouch, Robert E. Dewhurst, Winfried Hausner, Michael Thomm, and E. Peter Geiduschek (2003). "Activation of archaeal transcription by recruitment of the TATA-binding protein" (<http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=154304>). *Proceedings of the National Academy of Sciences of the United States of America* **100** (9): 5097. doi:10.1073/pnas.0837150100. PMID 12692306. PMC 154304.
- [7] Goldman, R.; Ebright, H.; Nickels, E. (May 2009). "Direct detection of abortive RNA transcripts in vivo" (<http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=2718712>). *Science (New York, N.Y.)* **324** (5929): 927–928. doi:10.1126/science.1169237. ISSN 0036-8075. PMID 19443781. PMC 2718712.
- [8] Dvir, A (Sep 2002). "Promoter escape by RNA polymerase II". *Biochimica et biophysica acta* **1577** (2): 208–223. ISSN 0006-3002. PMID 12213653.
- [9] Raj, A. and van Oudenaarden, A. (2008). Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* 135, 216-26.
- [10] "Chemistry 2006" (http://nobelprize.org/nobel_prizes/chemistry/laureates/2006/). *Nobel Foundation*. . Retrieved 2007-03-29.
- [11] <http://cmol.nbi.dk/models/dynamtrans/dynamtrans.html>
- [12] <http://cmol.nbi.dk/>
- [13] <http://cmol.nbi.dk/models/dna/rnap.html>
- [14] <http://highered.mcgraw-hill.com/sites/dl/free/0072437316/120060/ravenanimation.html>

DNA Transfer

In molecular biology, **transformation** is the genetic alteration of a cell resulting from the uptake, genomic incorporation, and expression of environmental genetic material (DNA).^[1] Transformation occurs most commonly in bacteria, both naturally and artificially, and refers to DNA taken up from the environment through their cell wall. Bacteria that are capable of being transformed are called competent. New genetic material can also be transferred to cells through conjugation or transduction. Conjugation involves cell-to-cell contact between two different bacterial cells, with the DNA being transferred from one bacterial cell to the other. In transduction, viruses called bacteriophages inject the foreign DNA into their host. Introduction of foreign DNA into eukaryotic cells is usually called "transfection".^[2] Transformation is also used to describe the insertion of new genetic material into nonbacterial cells including animal and plant cells.

History

Transformation was first demonstrated in 1928 by Frederick Griffith, an English bacteriologist searching for a vaccine against bacterial pneumonia. Griffith discovered that a harmless strain of *Streptococcus pneumoniae* could be made virulent after being exposed to heat-killed virulent strains. Griffith hypothesized that some "transforming factor" from the heat-killed strain was responsible for making the harmless strain virulent. In 1944 this "transforming factor" was identified as being genetic by Oswald Avery, Colin MacLeod, and Maclyn McCarty. They isolated DNA from virulent strain of *S. pneumoniae* and using just this DNA were able to make a harmless strain virulent. They called this uptake and incorporation of DNA by bacteria "transformation." See Avery-MacLeod-McCarty experiment.

The results of Avery *et al.*'s experiments were at first sceptically received by the scientific community, and it was not until the development of genetic markers and the discovery of other methods of genetic transfer (conjugation in 1947 and transduction in 1953) by Joshua Lederberg that Avery's experiments were accepted.^[3] Transformation did not become routine procedure in scientific laboratories until 1972, when Stanley Cohen, Annie Chang and Leslie Hsu successfully transformed *Escherichia coli* by treating the bacteria with calcium chloride.^[4] This created an efficient and convenient procedure for transforming DNA into bacteria and opened the way for molecular cloning in biotechnology and research.

Transformation using electroporation was developed in the late 1980s, increasing the efficiency and number of bacterial strains that could be transformed.^[5] Transformation of other animal and plant cells were investigated, with the first transgenic mouse being created by injecting genetic material that included a rat growth hormone gene into a mouse embryo in 1982.^[6] In 1907 a bacterium that caused plant tumors, *Agrobacterium tumefaciens*, was discovered and in the early 1970s the Tumour Inducing agent was found to be a DNA plasmid, called the Ti plasmid.^[7] By removing the genes in the plasmid that caused the cancer and adding in novel genes, researchers were able to infect plants with *A. tumefaciens* and let the bacteria insert their chosen DNA into the plants genome. Not all plant cells are susceptible to infection by *A. tumefaciens* so other methods were developed, including electroporation and micro-injection.^[8] Particle bombardment was made possible with the invention of the Biolistic Particle Delivery System (gene gun) by John Sanford in 1990.^[9]

Mechanisms

Bacteria

Bacteria transformation may be referred to as a stable genetic change brought about by taking up naked DNA (DNA without associated cells or proteins), and competence refers to the state of being able to take up exogenous DNA from the environment. Two different forms of competence should be distinguished: natural and artificial.

Natural competence

Some bacteria (around 1% of all species) are naturally capable of taking up DNA under laboratory conditions; many more may be able to take it up in their natural environments. Such species carry sets of genes specifying the cause of the machinery for bringing DNA across the cell's membrane or membranes.^[10]

Artificial competence

Artificial competence is not encoded in the cell's genes. Instead it is induced by laboratory procedures in which cells are passively made permeable to DNA, using conditions that do not normally occur in nature.^[11]

Calcium chloride transformation is a method of promoting competence. Chilling cells in the presence of divalent cations such as Ca^{2+} (in CaCl_2) prepares the cell membrane to become permeable to plasmid DNA. Cells are incubated on ice with the DNA and then briefly heat shocked (e.g. 42 °C for 30–120 seconds), which causes the DNA to enter the cell. This method works very well for circular plasmid DNAs. An excellent preparation of competent cells will give $\sim 10^8$ colonies per microgram of plasmid. A poor preparation will be about $10^4/\mu\text{g}$ or less. Good non-commercial preps should give 10^5 to 10^6 transformants per microgram of plasmid.

The method usually does not work well for linear molecules such as fragments of chromosomal DNA, probably because exonuclease enzymes in the cell rapidly degrade linear DNA. However, cells that are naturally competent are usually transformed more efficiently with linear DNA than with plasmids.

Electroporation is another way to make holes in bacterial (and other) cells, by briefly shocking them with an electric field of 10–20kV/cm. Plasmid DNA can enter the cell through these holes. This method is amenable to use with large plasmid DNA.^[12] Natural membrane-repair mechanisms will rapidly close these holes after the shock.

Plasmid transformation

In order to persist and be stably maintained in the cell, a plasmid DNA molecule must contain an origin of replication, which allows it to be replicated in the cell independently of the chromosome. Because transformation usually produces a mixture of rare transformed cells and abundant non-transformed cells, a method is needed to identify the cells that have acquired the plasmid. Plasmids used in transformation experiments will usually also contain a gene giving resistance to an antibiotic that the intended recipient strain of bacteria is sensitive to. Cells able to grow on media containing this antibiotic will have been transformed by the plasmid, as cells lacking the plasmid will be unable to grow.

Another marker, used for identifying *E. coli* cells that have acquired recombinant plasmids, is the *lacZ* gene, which codes for β -galactosidase. Because β -galactosidase is a homo-tetramer, with each monomer made up of one *lacZ*- α and one *lacZ*- ω protein, if only one of the two requisite proteins is expressed in the resulting cell, no functional enzyme will be formed. Thus, if a strain of *E. coli* without *lacZ*- α in its genome is transformed using a plasmid containing the missing gene fragment, transformed cells will produce β -galactosidase, while untransformed cells will not, as they are only able to produce the omega half of the monomer. In this type of transformation, the polylinker region of the plasmid lies in the *lacZ*- α gene fragment, meaning that successfully produced recombinant plasmids will have the desired gene inserted somewhere within *lacZ*- α . When this disrupted gene fragment is expressed by *E. coli*, no usable *lacZ*- α protein is produced, and therefore no usable β -galactosidase is formed. When grown on media containing the colorless, modified galactose sugar X-gal, colonies that are able to metabolize the substrate (and that

have therefore been transformed, but not by recombinant plasmids) will appear blue in color; colonies that are not able to metabolize the substrate (and that have therefore been transformed by recombinant plasmids) will appear white.

Plants

A number of mechanisms are available to transfer DNA into plant cells:

- *Agrobacterium* mediated transformation is the easiest and most simple plant transformation. Plant tissue (often leaves) is cut into small pieces, e.g. 10x10mm, and soaked for 10 minutes in a fluid containing suspended *Agrobacterium*. Some cells along the cut will be transformed by the bacterium, that inserts its DNA into the cell. Placed on selectable rooting and shooting media, the plants will regrow. Some plants species can be transformed just by dipping the flowers into suspension of *Agrobacterium* and then planting the seeds in a selective medium. Unfortunately, many plants are not transformable by this method.
- Particle bombardment: Particles of gold or tungsten are coated with DNA and then shot into young plant cells or plant embryos. Some genetic material will stay in the cells and transform them. This method also allows transformation of plant plastids. The transformation efficiency is lower than in agrobacterial mediated transformation, but most plants can be transformed with this method.
- Electroporation: make transient holes in cell membranes using electric shock; this allows DNA to enter as described above for Bacteria.
- Viral transformation (transduction): Package the desired genetic material into a suitable plant virus and allow this modified virus to infect the plant. If the genetic material is DNA, it can recombine with the chromosomes to produce transformant cells. However genomes of most plant viruses consist of single stranded RNA which replicates in the cytoplasm of infected cell. For such genomes this method is a form of transfection and not a real transformation, since the inserted genes never reach the nucleus of the cell and do not integrate into the host genome. The progeny of the infected plants is virus free and also free of the inserted gene.

Animals

Introduction of DNA into animal cells is usually called transfection, and is discussed in the corresponding article.

External links

- Bacterial Transformation ^[13] (a Flash Animation)

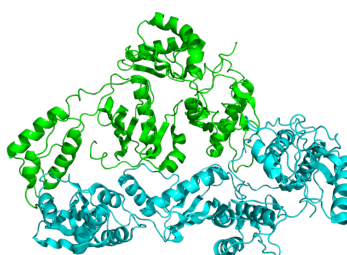
References

- [1] *bacterial transformation* (http://www.mercksource.com/pp/us/cns/cns_hl_dorlands_split.jsp?pg=/ppdocs/us/common/dorlands/dorland/eight/000110198.htm) at Dorland's Medical Dictionary
- [2] Alberts, Bruce; *et al.* (2002). *Molecular Biology of the Cell*. New York: Garland Science. p. G:35. ISBN 9780815340720.
- [3] Lederberg, Joshua (1994). *The Transformation of Genetics by DNA: An Anniversary Celebration of AVERY, MACLEOD and MCCARTY(1944) in Anecdotal, Historical and Critical Commentaries on Genetics* (<http://www.ncbi.nlm.nih.gov/pubmed/8150273>). The Rockefeller University, New York, New York 10021-6399. .
- [4] Cohen, Stanley; Chang, Annie and Hsu, Leslie (1972). "Nonchromosomal Antibiotic Resistance in Bacteria: Genetic Transformation of *Escherichia coli* by R-Factor DNA" (<http://www.pnas.org/content/69/8/2110.abstract>). *Proceedings of the National Academy of Sciences*. .
- [5] Wirth, Reinhard; Friesenegger, Anita and Fiedlerand, Stefan (1989). "Transformation of various species of gram-negative bacteria belonging to 11 different genera by electroporation" (<http://www.springerlink.com/content/xx826w544343jt8l/>). *Molecular and General Genetics MGG*. .
- [6] Palmiter, Richard; Ralph L. Brinster, Robert E. Hammer, Myrna E. Trumbauer, Michael G. Rosenfeld, Neal C. Birnberg & Ronald M. Evans (1982). "Dramatic growth of mice that develop from eggs microinjected with metallothionein–growth hormone fusion genes" (<http://www.nature.com/nature/journal/v300/n5893/abs/300611a0.html>). *Nature*. .
- [7] Nester, Eugene. "Agrobacterium: The Natural Genetic Engineer (100 Years Later)" (<http://www.apsnet.org/online/feature/Agrobacterium/>). . Retrieved 28 January 2010.

- [8] Peters, Pamela. "Transforming Plants - Basic Genetic Engineering Techniques" (http://www.accessexcellence.org/RC/AB/BA/Transforming_Plants.php). . Retrieved 28 January 2010.
- [9] Voiland, Michael; McCandless, Linda. "DEVELOPMENT OF THE "GENE GUN" AT CORNELL" (<http://www.nysaes.cornell.edu/pubs/press/1999/genegun.html>). . Retrieved 28th January 2010.
- [10] Chen I, Dubnau D (2004). "DNA uptake during bacterial transformation". *Nat. Rev. Microbiol.* **2** (3): 241–9. doi:10.1038/nrmicro844. PMID 15083159.
- [11] Large-volume transformation with high-throughput efficiency chemically competent cells. *Focus* 20:2 (1998).
- [12] Transformation efficiency of *E. coli* electroporated with large plasmid DNA. *Focus* 20:3 (1998).
- [13] <http://www.1lecture.com/Microbiology/Bacterial%20Transformation/index.html>

Reverse transcriptase

Reverse transcriptase (RNA-dependent DNA polymerase)



Crystallographic structure of HIV reverse transcriptase where the P51 subunit is colored green and the P66 subunit is colored cyan.^[1]

Identifiers

Symbol	RVT_1
Pfam	PF00078 ^[2]
InterPro	IPR000477 ^[3]
PROSITE	PS50878 ^[4]
SCOP	1hmv ^[5]

Available PDB structures:

1bqm^[6], 1bqn^[7], 1c0t^[8], 1c0u^[9], 1c1b^[10], 1c1c^[11], 1d0e^[12], 1d1u^[13], 1dlo^[14], 1dtq^[15], 1dtu^[16], 1eet^[17], 1ep4^[18], 1fk9^[19], 1fko^[20], 1fkp^[21], 1har^[22], 1hmv^[23], 1hni^[24], 1hmv^[25], 1hpz^[26], 1hqe^[27], 1hqu^[28], 1hvu^[29], 1hys^[30], 1i6j^[31], 1ikv^[32], 1ikw^[33], 1ikx^[34], 1iky^[35], 1j5o^[36], 1jkh^[37], 1jla^[38], 1jlb^[39], 1jlc^[40], 1jle^[41], 1jlf^[42], 1jlg^[43], 1jlg^[44], 1klm^[45], 1lw0^[46], 1lw2^[47], 1lwc^[48], 1lwe^[49], 1lwf^[50], 1mml^[51], 1mu2^[52], 1n4l^[53], 1n5y^[54], 1n6q^[55], 1nnd^[56], 1qai^[57], 1qaj^[58], 1qe1^[59], 1r0a^[60], 1rev^[61], 1rt1^[62], 1rt2^[63], 1rt3^[64], 1rt4^[65], 1rt5^[66], 1rt6^[67], 1rt7^[68], 1rtd^[69], 1rth^[70], 1rti^[71], 1rtj^[72], 1rw3^[73], 1s1t^[74], 1s1u^[75], 1s1v^[76], 1s1w^[77], 1s1x^[78], 1s6p^[79], 1s6q^[80], 1s9e^[81], 1s9g^[82], 1suq^[83], 1sv5^[84], 1t03^[85], 1t05^[86], 1tk1^[87], 1tkx^[88], 1tkz^[89], 1tl1^[90], 1tl3^[91], 1tv6^[92], 1tvr^[93], 1uwb^[94], 1vrt^[95], 1vru^[96], 1ztt^[97], 1ztw^[98], 2b5j^[99], 2b6a^[100], 2ban^[101], 2be2^[102], 2fjv^[103], 2fjw^[104], 2fjx^[105], 2fvp^[106], 2fvq^[107], 2fvr^[108], 2fvs^[109], 2hmi^[110], 2hnd^[111], 2hny^[112], 2hnz^[113], 2i5j^[114], 2iaj^[115], 2ic3^[116], 2opp^[117], 2opq^[118], 2opr^[119], 2ops^[120], 2r2r^[121], 2r2s^[122], 2r2t^[123], 2r2u^[124], 2rf2^[125], 2rki^[126], 2vg5^[127], 2vg6^[128], 2vg7^[129], 2zd1^[130], 2ze2^[131], 3bgr^[132], 3c6t^[133], 3c6u^[134], 3di6^[135], 3dle^[136], 3dlg^[137], 3dlk^[138], 3dm2^[139], 3dok^[140], 3dol^[141], 3drp^[142], 3drr^[143], 3drs^[144], 3hvt^[145]

RNA-directed DNA polymerase							
Identifiers							
EC number	2.7.7.49 ^[146]						
CAS number	9068-38-6 ^[147]						
IntEnz	IntEnz view ^[148]						
BRENDA	BRENDA entry ^[149]						
ExPASy	NiceZyme view ^[150]						
KEGG	KEGG entry ^[151]						
MetaCyc	metabolic pathway ^[152]						
PRIAM	profile ^[153]						
PDB	structures ^[154]						
Gene Ontology	AmiGO ^[155] / EGO ^[156]						
<table> <tr> <th colspan="2">Search</th></tr> <tr> <td>PMC</td><td>articles ^[157]</td></tr> <tr> <td>PubMed</td><td>articles ^[158]</td></tr> </table>		Search		PMC	articles ^[157]	PubMed	articles ^[158]
Search							
PMC	articles ^[157]						
PubMed	articles ^[158]						

In the fields of molecular biology and biochemistry, a **reverse transcriptase**, also known as **RNA-dependent DNA polymerase**, is a DNA polymerase enzyme that transcribes single-stranded RNA into double-stranded DNA. It also helps in the formation of a double helix DNA once the RNA has been reverse transcribed into a single strand cDNA. Normal transcription involves the synthesis of RNA from DNA; hence, reverse transcription is the *reverse* of this.

Reverse transcriptase was discovered by Howard Temin at the University of Wisconsin–Madison, and independently by David Baltimore in 1970 at MIT. The two shared the 1975 Nobel Prize in Physiology or Medicine with Renato Dulbecco for their discovery.

Well studied reverse transcriptases include:

- HIV-1 reverse transcriptase from human immunodeficiency virus type 1 (PDB 1HMY ^[159])
- M-MLV reverse transcriptase from the Moloney murine leukemia virus
- AMV reverse transcriptase from the avian myeloblastosis virus
- Telomerase reverse transcriptase that maintains the telomeres of eukaryotic chromosomes

Function in viruses

The enzyme is encoded and used by reverse-transcribing viruses, which use the enzyme during the process of replication. Reverse-transcribing RNA viruses, such as retroviruses, use the enzyme to reverse-transcribe their RNA genomes into DNA, which is then integrated into the host genome and replicated along with it. Reverse-transcribing DNA viruses, such as the hepadnaviruses, can allow RNA to serve as a template in assembling, and making DNA strands. HIV infects humans with the use of this enzyme. Without reverse transcriptase, the viral genome would not be able to incorporate into the host cell, resulting in the failure of the ability to replicate. Unlike bacteria, retroviruses use preexisting host-encoded transfer RNAs as primers.

Process of reverse transcription

Reverse transcriptase creates single stranded DNA from an RNA template.

In virus species with reverse transcriptase lacking DNA-dependent DNA polymerase activity, creation of double-stranded DNA can possibly be done by host-encoded DNA polymerase δ , mistaking the viral DNA-RNA for a primer and synthesizing a double-stranded DNA by similar mechanism as in primer removal, where the newly synthesized DNA displaces the original RNA template.

The process of reverse transcription is extremely error-prone and it is during this step that mutations may occur. Such mutations may cause drug resistance.

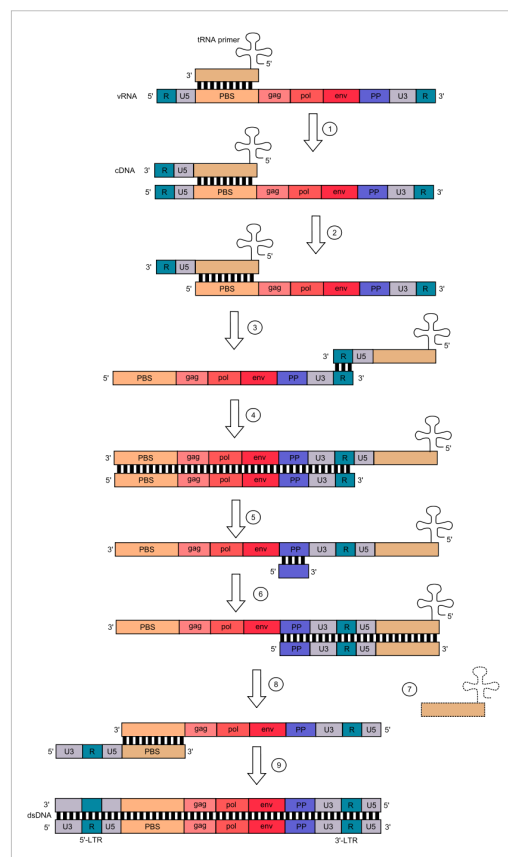
Process in class VI viruses

Class VI viruses ssRNA-RT, also called the retroviruses are RNA reverse transcribing viruses with a DNA intermediate. Their genomes consist of two molecules of positive sense single stranded RNA with a 5' cap and 3' polyadenylated tail. Examples of retroviruses include *Human Immunodeficiency Virus* (HIV) and *Human T-Lymphotropic virus* (HTLV). Creation of double-stranded DNA occurs in the cytosol^[160] as a series of steps:

1. A specific cellular tRNA acts as a primer and hybridizes to a complementary part of the virus genome called the primer binding site or PBS
2. Complementary DNA then binds to the U5 (non-coding region) and R region (a direct repeat found at both ends of the RNA molecule) of the viral RNA
3. A domain on the reverse transcriptase enzyme called RNase H degrades the 5' end of the RNA which removes the U5 and R region
4. The primer then 'jumps' to the 3' end of the viral genome and the newly synthesised DNA strands hybridizes to the complementary R region on the RNA
5. The first strand of complementary DNA (cDNA) is extended and the majority of viral RNA is degraded by RNase H
6. Once the strand is completed, second strand synthesis is initiated from the viral RNA
7. There is then another 'jump' where the PBS from the second strand hybridizes with the complementary PBS on the first strand
8. Both strands are extended further and can be incorporated into the hosts genome by the enzyme integrase

Creation of double-stranded DNA also involves *strand transfer*, in which there is a translocation of short DNA product from initial RNA dependent DNA synthesis to acceptor template regions at the other end of the genome, which are later reached and processed by the reverse transcriptase for its DNA-dependent DNA activity.^[161]

Retroviral RNA is arranged in 5' terminus to 3' terminus. The site where the primer is annealed to viral RNA is called the primer-binding site (PBS). The RNA 5' end to the PBS site is called U5, and the RNA 3' end to the PBS is called the leader. The tRNA primer is unwound between 14 and 22 nucleotides and forms a base-paired duplex with the viral RNA at PBS. That PBS located near the 5' terminus of viral RNA is unusual because reverse transcriptase synthesizes DNA from 3' end of the primer in the 5' to 3' direction. Therefore, the primer and reverse transcriptase must be relocated to 3' end of viral RNA. In order to accomplish this reposition, multiple steps and various enzymes



including DNA polymerase, ribonuclease H(RNase H) and polynucleotide unwinding are needed.^[162]

The HIV reverse transcriptase also has ribonuclease activity that degrades the viral RNA during the synthesis of cDNA, as well as DNA-dependent DNA polymerase activity that copies the sense cDNA strand into an *antisense* DNA to form a double-stranded viral DNA intermediate (vDNA).^[163]

In eukaryotes

Self-replicating stretches of eukaryotic genomes known as retrotransposons utilize reverse transcriptase to move from one position in the genome to another via a RNA intermediate. They are found abundantly in the genomes of plants and animals. Telomerase is another reverse transcriptase found in many eukaryotes, including humans, which carries its own RNA template; this RNA is used as a template for DNA replication.^{[164] [165]}

In prokaryotes

Reverse transcriptases are also found in bacterial Retron msr RNAs, distinct sequences which code for reverse transcriptase, and are used in the synthesis of msDNA. In order to initiate synthesis of DNA, a primer is needed. In bacteria, the primer is synthesized during replication.^[166]

Structure

Reverse transcriptase enzymes include an RNA-dependent DNA polymerase and a DNA-dependent DNA polymerase, which work together to perform transcription. In addition to the transcription function, retroviral reverse transcriptases have a domain belonging to the RNase H family which is vital to their replication.

Replication fidelity

There are three different replication systems during the life cycle of a retrovirus. First of all, the reverse transcriptase synthesize viral DNA from viral RNA, and then from newly made complementary DNA strand. The second replication process occurs when host cellular DNA polymerase replicates the integrated viral DNA. Lastly, RNA polymerase II transcribes the proviral DNA into RNA which will be packed into virions. Therefore, mutation can occur during one or all of these replication steps.^[167]

Reverse transcriptase has a high error rate when transcribing RNA into DNA since, unlike any other DNA polymerases, it has no proofreading ability. This high error rate allows mutations to accumulate at an accelerated rate relative to proofread forms of replication. The commercially available reverse transcriptases produced by Promega are quoted by their manuals as having error rates in the range of 1 in 17,000 bases for AMV and 1 in 30,000 bases for M-MLV^[168]

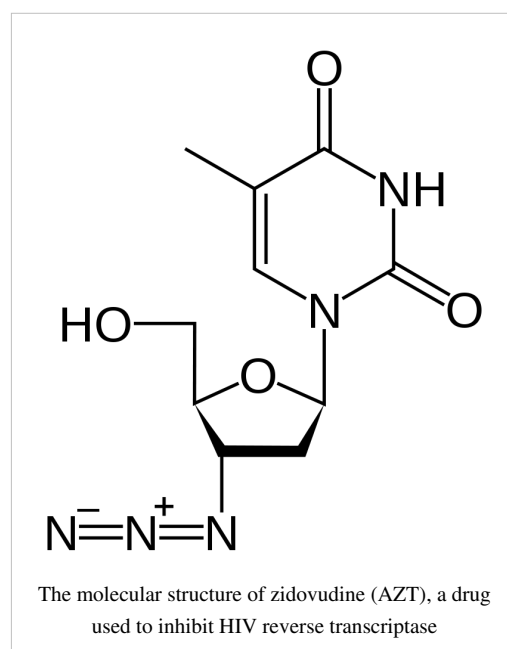
Applications

Antiviral drugs

As HIV uses reverse transcriptase to copy its genetic material and generate new viruses (part of a retrovirus proliferation circle), specific drugs have been designed to disrupt the process and thereby suppress its growth. Collectively, these drugs are known as reverse transcriptase inhibitors and include the nucleoside and nucleotide analogues zidovudine (trade name Retrovir), lamivudine (Epivir) and tenofovir (Viread), as well as non-nucleoside inhibitors, such as nevirapine (Viramune).

Molecular biology

Reverse transcriptase is commonly used in research to apply the polymerase chain reaction technique to RNA in a technique called reverse transcription polymerase chain reaction (RT-PCR). The classical PCR technique can be applied only to DNA strands, but, with the help of reverse transcriptase, RNA can be transcribed into DNA, thus making PCR analysis of RNA molecules possible. Reverse transcriptase is used also to create cDNA libraries from mRNA. The commercial availability of reverse transcriptase greatly improved knowledge in the area of molecular biology, as, along with other enzymes, it allowed scientists to clone, sequence, and characterise DNA.



History

The idea of reverse transcription was very unpopular at first as it contradicted the central dogma of molecular biology which states that DNA is transcribed into RNA which is then translated into proteins. However, in 1970 when the scientists Howard Temin and David Baltimore both independently discovered the enzyme responsible for reverse transcription, named reverse transcriptase, the possibility that genetic information could be passed on in this manner was finally accepted.

See also

- cDNA library
- DNA polymerase
- msDNA
- Reverse transcribing virus
- RNA polymerase
- Telomerase
- Retrotransposon marker

External links

- MeSH *RNA+Transcriptase* ^[169]
- animation of reverse transcriptase action and three reverse transcriptase inhibitors ^[170]
- Molecule of the month ^[171] (September 2002) at the Protein Data Bank
- HIV Replication 3D Medical Animation. (Nov 2008). Video by Boehringer Ingelheim. ^[172]

References

- [1] PDB 1HMY (<http://www.rcsb.org/pdb/explore/explore.do?structureId=1HMY>); Rodgers DW, Gamblin SJ, Harris BA, Ray S, Culp JS, Hellmig B, Woolf DJ, Debouck C, Harrison SC (February 1995). "The structure of unliganded reverse transcriptase from the human immunodeficiency virus type 1" (<http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=42671>). *Proc. Natl. Acad. Sci. U.S.A.* **92** (4): 1222–6. PMID 7532306. PMC 42671.
- [2] <http://pfam.sanger.ac.uk/family?acc=PF00078>
- [3] <http://www.ebi.ac.uk/interpro/DisplayIproEntry?ac=IPR000477>
- [4] <http://www.expasy.org/cgi-bin/prosite-search-ac?PS50878>
- [5] <http://scop.mrc-lmb.cam.ac.uk/scop/search.cgi?lev=fa;&pdb=1hmv>
- [6] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1bqm>
- [7] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1bqn>
- [8] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1c0t>
- [9] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1c0u>
- [10] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1c1b>
- [11] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1c1c>
- [12] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1d0e>
- [13] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1d1u>
- [14] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1dlo>
- [15] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1dtq>
- [16] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1dtl>
- [17] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1eet>
- [18] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1ep4>
- [19] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1fk9>
- [20] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1fko>
- [21] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1fkp>
- [22] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1har>
- [23] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1hmv>
- [24] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1hni>
- [25] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1hmv>
- [26] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1hpx>
- [27] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1hqe>
- [28] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1hqu>
- [29] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1hvu>
- [30] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1hys>
- [31] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1i6j>
- [32] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1ikv>
- [33] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1ikw>
- [34] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1ikx>
- [35] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1iky>
- [36] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1j5o>
- [37] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1jkh>
- [38] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1jla>
- [39] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1jlb>
- [40] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1jlc>
- [41] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1jle>
- [42] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1jlf>
- [43] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1jlg>
- [44] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1jll>
- [45] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1klm>
- [46] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1lw0>

- [47] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1lw2>
- [48] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1lwc>
- [49] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1lwe>
- [50] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1lwf>
- [51] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1mml>
- [52] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1mu2>
- [53] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1n4l>
- [54] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1n5y>
- [55] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1n6q>
- [56] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1nnd>
- [57] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1qai>
- [58] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1qaj>
- [59] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1qe1>
- [60] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1r0a>
- [61] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1rev>
- [62] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1rt1>
- [63] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1rt2>
- [64] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1rt3>
- [65] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1rt4>
- [66] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1rt5>
- [67] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1rt6>
- [68] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1rt7>
- [69] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1rtd>
- [70] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1rth>
- [71] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1rti>
- [72] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1rtj>
- [73] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1rw3>
- [74] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1s1t>
- [75] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1s1u>
- [76] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1s1v>
- [77] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1s1w>
- [78] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1s1x>
- [79] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1s6p>
- [80] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1s6q>
- [81] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1s9e>
- [82] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1s9g>
- [83] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1suq>
- [84] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1sv5>
- [85] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1t03>
- [86] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1t05>
- [87] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1tkl>
- [88] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1tkx>
- [89] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1tkz>
- [90] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1tl1>
- [91] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1tl3>
- [92] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1tv6>
- [93] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1tvr>
- [94] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1uwb>
- [95] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1vrt>
- [96] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1vru>
- [97] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1ztt>
- [98] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1ztw>
- [99] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=2b5j>
- [100] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=2b6a>
- [101] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=2ban>
- [102] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=2be2>
- [103] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=2fjv>
- [104] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=2fjw>
- [105] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=2fjx>

- [106] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=2fvp>
- [107] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=2fvq>
- [108] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=2fvr>
- [109] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=2fvs>
- [110] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=2hmi>
- [111] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=2hnd>
- [112] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=2hny>
- [113] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=2hnz>
- [114] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=2i5j>
- [115] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=2iaj>
- [116] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=2ic3>
- [117] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=2opp>
- [118] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=2opq>
- [119] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=2opr>
- [120] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=2ops>
- [121] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=2r2r>
- [122] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=2r2s>
- [123] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=2r2t>
- [124] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=2r2u>
- [125] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=2rf2>
- [126] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=2rki>
- [127] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=2vg5>
- [128] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=2vg6>
- [129] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=2vg7>
- [130] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=2zd1>
- [131] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=2ze2>
- [132] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=3bgr>
- [133] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=3c6t>
- [134] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=3c6u>
- [135] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=3di6>
- [136] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=3dle>
- [137] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=3dlg>
- [138] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=3dlk>
- [139] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=3dm2>
- [140] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=3dok>
- [141] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=3dol>
- [142] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=3drp>
- [143] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=3drr>
- [144] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=3drs>
- [145] <http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=3hvt>
- [146] <http://www.chem.qmul.ac.uk/iubmb/enzyme/EC2/7/7/49.html>
- [147] <http://toolserver.org/~magnus/cas.php?language=en&cas=9068-38-6&title=>
- [148] <http://www.ebi.ac.uk/intenz/query?cmd=SearchEC&ec=2.7.7.49>
- [149] http://www.brenda-enzymes.org/php/result_flat.php4?ecno=2.7.7.49
- [150] <http://www.expasy.org/enzyme/2.7.7.49>
- [151] http://www.genome.ad.jp/dbget-bin/www_bget?enzyme+2.7.7.49
- [152] <http://biocyc.org/META/substring-search?type=NIL&object=2.7.7.49>
- [153] http://bioinfo.genopole-toulouse.prd.fr/priam/cgi-bin/PRIAM_profiles_CurrentRelease.pl?EC=2.7.7.49
- [154] http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/enzymes/GetPage.pl?ec_number=2.7.7.49
- [155] <http://amigo.geneontology.org/cgi-bin/amigo/go.cgi?query=GO:0003964&view=details>
- [156] <http://www.ebi.ac.uk/ego/DisplayGoTerm?id=GO:0003964&format=normal>
- [157] <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pubmed&term=2.7.7.49%5BEC/RN%20Number%5D%20AND%20pubmed%20pmc%20local%5Bs%5D>
- [158] <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pubmed&term=2.7.7.49%5BEC/RN%20Number%5D>
- [159] <http://www.rcsb.org/pdb/explore/explore.do?structureId=1HMY>
- [160] Bio-Medicine.org - Retrovirus (<http://www.bio-medicine.org/biology-definition/Retrovirus/>) Retrieved on 17 Feb, 2009
- [161] Telesnitsky, A., Goff, S.P. (1993). "Strong-stop strand transfer during reverse transcription". in Skalka, M. A., Goff, S.P. *Reverse transcriptase* (1st ed.). New York: Cold Spring Harbor. p. 49. ISBN 0-87969-382-7.

- [162] Bernstein, A.; Weiss, Robin; Tooze, John (1985). "RNA tumor viruses". *Molecular Biology of Tumor Viruses* (2nd ed.). Cold Spring Harbor, N.Y: Cold Spring Harbor Laboratory.
- [163] Doc Kaiser's Microbiology Home Page > IV. VIRUSES > F. ANIMAL VIRUS LIFE CYCLES > 3. The Life Cycle of HIV (<http://student.ccbcmd.edu/courses/bio141/lecguide/unit3/viruses/hivlc.html>) Community College of Baltimore County. Updated: Jan., 2008
- [164] Monty Krieger; Matthew P Scott; Matsudaira, Paul T.; Lodish, Harvey F.; Darnell, James E.; Lawrence Zipursky; Kaiser, Chris; Arnold Berk (2004). *Molecular cell biology* (5th ed.). New York: W.H. Freeman and CO. ISBN 0-7167-4366-3.
- [165] Witzany G (August 2008). "The Viral Origins of Telomeres and Telomerases and their Important Role in Eukaryogenesis and Genome Maintenance". *Biosemiotics* **1** (2): 191–206. doi:10.1007/s12304-008-9018-0.
- [166] Hurwitz J, Leis JP (January 1972). "RNA-dependent DNA polymerase activity of RNA tumor viruses. I. Directing influence of DNA in the reaction" (<http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=356270>). *J. Virol.* **9** (1): 116–29. PMID 4333538. PMC 356270.
- [167] Bbenek, K., Kunkel, A. T (1993). "The fidelity of retroviral reverse transcriptases". in Skalka, M. A., Goff, P. S.. *Reverse transcriptase*. New York: Cold Spring Harbor Laboratory Press. p. 85. ISBN 0-87969-382-7.
- [168] Promega kit instruction manual (1999) (http://www.promega.com/pnotes/71/7807_22/7807_22_core.pdf)
- [169] http://www.nlm.nih.gov/cgi/mesh/2009/MB_cgi?mode=&term=RNA+Transcriptase
- [170] http://www.tibotec.com/bgdisplay.jhtml?itemname=HIV_discovery&product=none&s=2
- [171] http://www.rcsb.org/pdb/static.do?p=education_discussion/molecule_of_the_month/pdb33_1.html
- [172] <http://www.youtube.com/watch?v=RO8MP3wMvqg>

DNA microarray

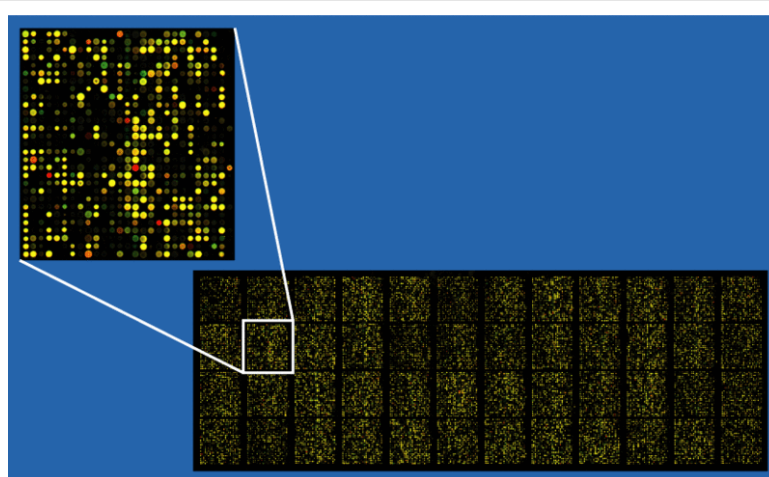
For terminology, see glossary below.

A **DNA microarray** is a multiplex technology used in molecular biology and in medicine. It consists of an arrayed series of thousands of microscopic spots of DNA oligonucleotides, called features, each containing picomoles (10^{-12} moles) of a specific DNA sequence, known as *probes* (or *reporters*). This can be a short section of a gene or other DNA element that are used to hybridize a cDNA or cRNA sample (called *target*) under high-stringency conditions.

Probe-target hybridization is usually detected and quantified by detection of fluorophore-, silver-, or chemiluminescence-labeled targets to determine relative abundance of nucleic acid sequences in the target. Since an array can contain tens of thousands of probes, a microarray experiment can accomplish many genetic tests in parallel. Therefore arrays have dramatically accelerated many types of investigation.

In standard microarrays, the probes are attached via surface engineering to a solid surface by a covalent bond to a chemical matrix (via epoxy-silane, amino-silane, lysine, polyacrylamide or others). The solid surface can be glass or a silicon chip, in which case they are commonly known as *gene chip* or colloquially *Affy chip* when an Affymetrix chip is used. Other microarray platforms, such as Illumina, use microscopic beads, instead of the large solid support. DNA arrays are different from other types of microarray only in that they either measure DNA or use DNA as part of its detection system.

DNA microarrays can be used to measure changes in expression levels, to detect single nucleotide polymorphisms (SNPs), to genotype or resequence mutant genomes (*see uses and types section*). Microarrays also differ in fabrication, workings, accuracy, efficiency, and cost (*see fabrication section*). Additional factors for microarray



Example of an approximately 40,000 probe spotted oligo microarray with enlarged inset to show detail.

experiments are the experimental design and the methods of analyzing the data (*see Bioinformatics section*).

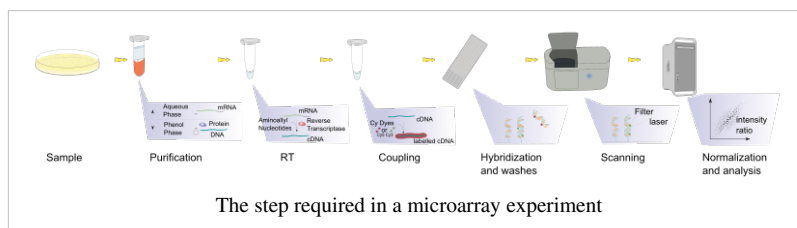
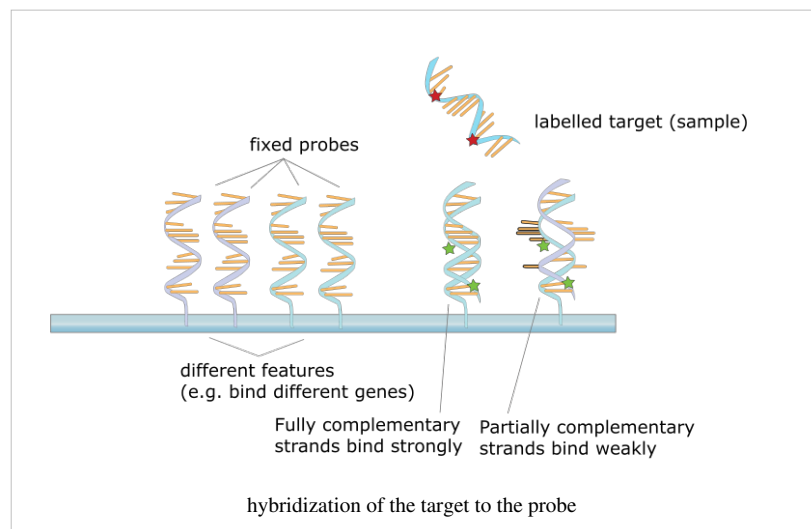
History

Microarray technology evolved from Southern blotting, where fragmented DNA is attached to a substrate and then probed with a known gene or fragment. The use of a collection of distinct DNAs in arrays for expression profiling was first described in 1987, and the arrayed DNAs were used to identify genes whose expression is modulated by interferon.^[1] These early gene arrays were made by spotting cDNAs onto filter paper with a pin-spotting device. The use of miniaturized microarrays for gene expression profiling was first reported in 1995,^[2] and a complete eukaryotic genome (*Saccharomyces cerevisiae*) on a microarray was published in 1997.^[3]

Principle

The core principle behind microarrays is hybridization between two DNA strands, the property of complementary nucleic acid sequences to specifically pair with each other by forming hydrogen bonds between complementary nucleotide base pairs. A high number of complementary base pairs in a nucleotide sequence means tighter non-covalent bonding between the two strands. After washing off of non-specific bonding sequences, only strongly paired strands will remain hybridized. So fluorescently labelled

target sequences that bind to a probe sequence generate a signal that depends on the strength of the hybridization determined by the number of paired bases, the hybridization conditions (such as temperature), and washing after hybridization. Total strength of the signal, from a spot (feature), depends upon the amount of target sample binding to the probes present on that spot. Microarrays use relative quantitation in which the intensity of a feature is compared to the intensity of the same feature under a different condition, and the identity of the feature is known by its position. An alternative to microarrays is serial analysis of gene expression, where the transcriptome is sequenced allowing an absolute measurement.



Uses and types

Many types of array exist and the broadest distinction is whether they are spatially arranged on a surface or on coded beads:

- The traditional solid-phase array is a collection of orderly microscopic "spots", called features, each with a specific probe attached to a solid surface, such as glass, plastic or silicon biochip (commonly known *gene chip*, *genome chip*, *DNA chip* or *gene array*). Thousands of them can be placed in known locations on a single DNA microarray.
- The alternative bead array is a collection of microscopic polystyrene beads, each with a specific probe and a ratio of two or more dyes, which do not interfere with the fluorescent dyes used on the target sequence.



DNA microarrays can be used to detect DNA (as in comparative genomic hybridization), or detect RNA (most commonly as cDNA after reverse transcription) that may or may not be translated into proteins. The process of measuring gene expression via cDNA is called expression analysis or expression profiling.

Applications include:

Application or technology	Synopsis
Gene expression profiling	In an mRNA or gene expression profiling experiment the expression levels of thousands of genes are simultaneously monitored to study the effects of certain treatments, diseases, and developmental stages on gene expression. For example, microarray-based gene expression profiling can be used to identify genes whose expression is changed in response to pathogens or other organisms by comparing gene expression in infected to that in uninfected cells or tissues. ^[4]
Comparative genomic hybridization	Assessing genome content in different cells or closely related organisms. ^{[5] [6]}
GeneID	Small microarrays to check IDs of organisms in food and feed (like GMO [7]), mycoplasmas in cell culture, or pathogens for disease detection, mostly combining PCR and microarray technology.
Chromatin immunoprecipitation on Chip	DNA sequences bound to a particular protein can be isolated by immunoprecipitating that protein (ChIP), these fragments can be then hybridized to a microarray (such as a tiling array) allowing the determination of protein binding site occupancy throughout the genome. Example protein to immunoprecipitate are histone modifications (H3K27me3, H3K4me2, H3K9me3, etc), Polycomb-group protein (PRC2:Suz12, PRC1:YY1) and trithorax-group protein (Ash1) to study the epigenetic landscape or RNA Polymerase II to study the transcription landscape.
DamID	Analogously to ChIP, genomic regions bound by a protein of interest can be isolated and used to probe a microarray to determine binding site occupancy. Unlike ChIP, DamID does not require antibodies but makes use of adenine methylation near the protein's binding sites to selectively amplify those regions, introduced by expressing minute amounts of protein of interest fused to bacterial DNA adenine methyltransferase.
SNP detection	Identifying single nucleotide polymorphism among alleles within or between populations. ^[8] Several applications of microarrays make use of SNP detection, including Genotyping, forensic analysis, measuring predisposition to disease, identifying drug-candidates, evaluating germline mutations in individuals or somatic mutations in cancers, assessing loss of heterozygosity, or genetic linkage analysis.
Alternative splicing detection	An <i>'exon junction array'</i> design uses probes specific to the expected or potential splice sites of predicted exons for a gene. It is of intermediate density, or coverage, to a typical gene expression array (with 1-3 probes per gene) and a genomic tiling array (with hundreds or thousands of probes per gene). It is used to assay the expression of alternative splice forms of a gene. Exon arrays have a different design, employing probes designed to detect each individual exon for known or predicted genes, and can be used for detecting different splicing isoforms.
Fusion genes microarray	A Fusion gene microarray can detect fusion transcripts, <i>e.g.</i> from cancer specimens. The principle behind this is building on the alternative splicing microarrays. The oligo design strategy enables combined measurements of chimeric transcript junctions with exon-wise measurements of individual fusion partners.

Tiling array	Genome tiling arrays consist of overlapping probes designed to densely represent a genomic region of interest, sometimes as large as an entire human chromosome. The purpose is to empirically detect expression of transcripts or alternatively splice forms which may not have been previously known or predicted.
--------------	--

Fabrication

Microarrays can be manufactured in different ways, depending on the number of probes under examination, costs, customization requirements, and the type of scientific question being asked. Arrays may have as few as 10 probes or up to 2.1 million (NimbleGen, Roche) micrometre-scale probes from commercial vendors.

Spotted vs. in situ synthesised arrays

Microarrays can be fabricated using a variety of technologies, including printing with fine-pointed pins onto glass slides, photolithography using pre-made masks, photolithography using dynamic micromirror devices, ink-jet printing,^[9] or electrochemistry on microelectrode arrays.

In *spotted microarrays*, the probes are oligonucleotides, cDNA or small fragments of PCR products that correspond to mRNAs. The probes are synthesized prior to deposition on the array surface and are then "spotted" onto glass. A common approach utilizes an array of fine pins or needles controlled by a robotic arm that is dipped into wells containing DNA probes and then depositing each probe at designated locations on the array surface. The resulting "grid" of probes represents the nucleic acid profiles of the prepared probes and is ready to receive complementary cDNA or cRNA "targets" derived from experimental or clinical samples. This technique is used by research scientists around the world to produce "in-house" printed microarrays from their own labs. These arrays may be easily customized for each experiment, because researchers can choose the probes and printing locations on the arrays, synthesize the probes in their own lab (or collaborating facility), and spot the arrays. They can then generate their own labeled samples for hybridization, hybridize the samples to the array, and finally scan the arrays with their own equipment. This provides a relatively low-cost microarray that may be customized for each study, and avoids the costs of purchasing often more expensive commercial arrays that may represent vast numbers of genes that are not of interest to the investigator. Publications exist which indicate in-house spotted microarrays may not provide the same level of sensitivity compared to commercial oligonucleotide arrays,^[10] possibly owing to the small batch sizes and reduced printing efficiencies when compared to industrial manufactures of oligo arrays.

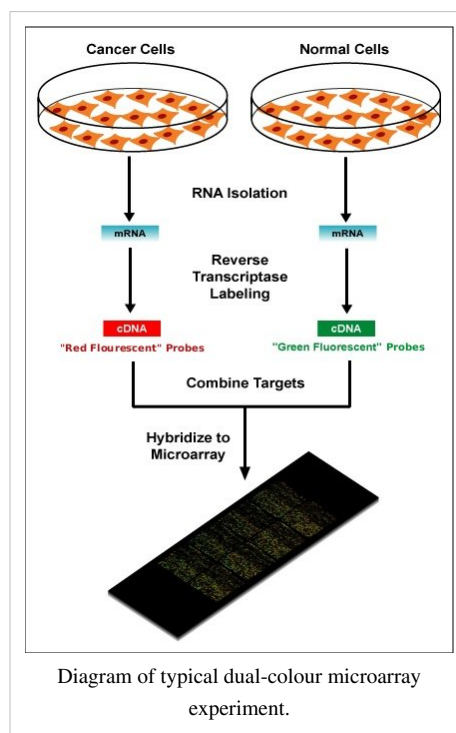
In *oligonucleotide microarrays*, the probes are short sequences designed to match parts of the sequence of known or predicted open reading frames. Although oligonucleotide probes are often used in "spotted" microarrays, the term "oligonucleotide array" most often refers to a specific technique of manufacturing. Oligonucleotide arrays are produced by printing short oligonucleotide sequences designed to represent a single gene or family of gene splice-variants by synthesizing this sequence directly onto the array surface instead of depositing intact sequences. Sequences may be longer (60-mer probes such as the Agilent design) or shorter (25-mer probes produced by Affymetrix) depending on the desired purpose; longer probes are more specific to individual target genes, shorter probes may be spotted in higher density across the array and are cheaper to manufacture. One technique used to produce oligonucleotide arrays include photolithographic synthesis (Agilent and Affymetrix) on a silica substrate where light and light-sensitive masking agents are used to "build" a sequence one nucleotide at a time across the entire array.^[11] Each applicable probe is selectively "unmasked" prior to bathing the array in a solution of a single nucleotide, then a masking reaction takes place and the next set of probes are unmasked in preparation for a different nucleotide exposure. After many repetitions, the sequences of every probe become fully constructed. More recently, Maskless Array Synthesis from NimbleGen Systems has combined flexibility with large numbers of probes.^[12]

Two-channel vs. one-channel detection

Two-color microarrays or *two-channel microarrays* are typically hybridized with cDNA prepared from two samples to be compared (e.g. diseased tissue versus healthy tissue) and that are labeled with two different fluorophores.^[13] Fluorescent dyes commonly used for cDNA labeling include Cy3, which has a fluorescence emission wavelength of 570 nm (corresponding to the green part of the light spectrum), and Cy5 with a fluorescence emission wavelength of 670 nm (corresponding to the red part of the light spectrum). The two Cy-labeled cDNA samples are mixed and hybridized to a single microarray that is then scanned in a microarray scanner to visualize fluorescence of the two fluorophores after excitation with a laser beam of a defined wavelength. Relative intensities of each fluorophore may then be used in ratio-based analysis to identify up-regulated and down-regulated genes.^[14]

Oligonucleotide microarrays often carry control probes designed to hybridize with RNA spike-ins. The degree of hybridization between the spike-ins and the control probes is used to normalize the hybridization measurements for the target probes. Although absolute levels of gene expression may be determined in the two-color array in rare instances, the relative differences in expression among different spots within a sample and between samples is the preferred method of data analysis for the two-color system. Examples of providers for such microarrays includes Agilent with their Dual-Mode platform, Eppendorf with their DualChip platform for colorimetric Silverquant labeling, and TeleChem International with Arrayit.

In *single-channel microarrays* or *one-color microarrays*, the arrays provide intensity data for each probe or probe set indicating a relative level of hybridization with the labeled target. However, they do not truly indicate abundance levels of a gene but rather relative abundance when compared to other samples or conditions when processed in the same experiment. Each RNA molecule encounters protocol and batch-specific bias during amplification, labeling, and hybridization phases of the experiment making comparisons between genes for the same microarray uninformative. The comparison of two conditions for the same gene requires two separate single-dye hybridizations. Several popular single-channel systems are the Affymetrix "Gene Chip", Illumina "Bead Chip", Agilent single-channel arrays, the Applied Microarrays "CodeLink" arrays, and the Eppendorf "DualChip & Silverquant". One strength of the single-dye system lies in the fact that an aberrant sample cannot affect the raw data derived from other samples, because each array chip is exposed to only one sample (as opposed to a two-color system in which a single low-quality sample may drastically impinge on overall data precision even if the other sample was of high quality). Another benefit is that data are more easily compared to arrays from different experiments so long as batch effects have been accounted for. A drawback to the one-color system is that, when compared to the two-color system, twice as many microarrays are needed to compare samples within an experiment.



Microarrays and bioinformatics

The advent of inexpensive microarray experiments created several specific bioinformatics challenges:

- the multiple levels of replication in experimental design (Experimental design)
- the number of platforms and independent groups and data format (Standardization)
- the treatment of the data (Statistical analysis)
- accuracy and precision (Relation between probe and gene)
- the sheer volume of data and the ability to share it (Data warehousing)

Experimental design

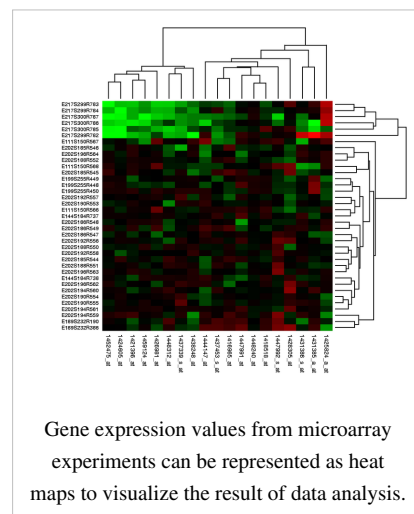
Due to the biological complexity of gene expression, the considerations of experimental design that are discussed in the expression profiling article are of critical importance if statistically and biologically valid conclusions are to be drawn from the data.

There are three main elements to consider when designing a microarray experiment. First, replication of the biological samples is essential for drawing conclusions from the experiment. Second, technical replicates (two RNA samples obtained from each experimental unit) help to ensure precision and allow for testing differences within treatment groups. The technical replicates may be two independent RNA extractions or two aliquots of the same extraction. Third, spots of each cDNA clone or oligonucleotide are present as replicates (at least duplicates) on the microarray slide, to provide a measure of technical precision in each hybridization. It is critical that information about the sample preparation and handling is discussed, in order to help identify the independent units in the experiment and to avoid inflated estimates of statistical significance.^[15]

Standardization

Microarray data is difficult to exchange due to the lack of standardization in platform fabrication, assay protocols, and analysis methods. This presents an interoperability problem in bioinformatics. Various grass-roots open-source projects are trying to ease the exchange and analysis of data produced with non-proprietary chips:

- For example, the "Minimum Information About a Microarray Experiment" (MIAME) checklist helps define the level of detail that should exist and is being adopted by many journals as a requirement for the submission of papers incorporating microarray results. But MIAME does not describe the format for the information, so while many formats can support the MIAME requirements, as of 2007 no format permits verification of complete semantic compliance.
- The "MicroArray Quality Control (MAQC) Project" is being conducted by the US Food and Drug Administration (FDA) to develop standards and quality control metrics which will eventually allow the use of MicroArray data in drug discovery, clinical practice and regulatory decision-making.^[16]
- The MGED Society has developed standards for the representation of gene expression experiment results and relevant annotations.



Statistical analysis

Microarray data sets are commonly very large, and analytical precision is influenced by a number of variables. Statistical challenges include taking into account effects of background noise and appropriate normalization of the data. Normalization methods may be suited to specific platforms and, in the case of commercial platforms, the analysis may be proprietary. Algorithms that affect statistical analysis include:

- Image analysis: gridding, spot recognition of the scanned image (segmentation algorithm), removal or marking of poor-quality and low-intensity features (called *flagging*).
- Data processing: background subtraction (based on global or local background), determination of spot intensities and intensity ratios, visualisation of data (e.g see MA plot), and log-transformation of ratios, global or local normalization of intensity ratios.
- Identification of statistically significant changes: t-test, ANOVA, Bayesian method or Mann–Whitney test methods tailored to microarray data sets, which take into account multiple comparisons.^[17] These methods assess statistical power based on the variation present in the data and the number of experimental replicates, and can help minimize Type I and type II errors in the analyses.^[18]
- Network-based methods: Statistical methods that take the underlying structure of gene networks into account, representing either associative or causative interactions or dependencies among gene products.^[19]

Microarray data may require further processing aimed at reducing the dimensionality of the data to aid comprehension and more focused analysis.^[20] Other methods permit analysis of data consisting of a low number of biological or technical replicates; for example, the Local Pooled Error (LPE) test pools standard deviations of genes with similar expression levels in an effort to compensate for insufficient replication.^[21]

Relation between probe and gene

The relation between a probe and the mRNA that it is expected to detect is problematic. On the one hand, some mRNAs may cross-hybridize probes in the array that are supposed to detect another mRNA. In addition, mRNAs may experience amplification bias that is sequence or molecule-specific. On the other hand, probes that are designed to detect the mRNA of a particular gene may be relying on genomic EST information that is incorrectly associated with that gene.

Data warehousing

Microarray data was found to be more useful when compared to other similar datasets. The sheer volume (in bytes), specialized formats (such as MIAME), and curation efforts associated with the datasets require specialized databases to store the data.

See also

- Cyanine dyes, such as Cy3 and Cy5, are commonly used fluorophores with microarrays
 - FatiGO
 - Full Genome Sequencing
 - Gene chip analysis
 - Microfluidics or lab-on-chip
 - Serial analysis of gene expression
 - Significance analysis of microarrays
 - Systems biology
-

Glossary

- An **Array** or **slide** is a collection of *features* spatially arranged in a two dimensional grid, arranged in columns and rows.
- **Block** or **subarray**: a group of spots, typically made in one print round; several subarrays/blocks form an array.
- **Case/control**: an experimental design paradigm especially suited to the two-colour array system, in which a condition chosen as control (such as healthy tissue or state) is compared to an altered condition (such as a diseased tissue or state).
- **Channel**: the fluorescence output recorded in the scanner for an individual fluorophore and can even be ultraviolet.
- **Dye flip** or **Dye swap** or **Fluor reversal**: reciprocal labelling of DNA targets with the two dyes to account for dye bias in experiments.
- **Scanner**: an instrument used to detect and quantify the intensity of fluorescence of spots on a microarray slide, by selectively exciting fluorophores with a laser and measuring the fluorescence with a filter (optics) photomultiplier system.
- **Spot** or **feature**: a small area on an array slide that contains picomoles of specific DNA samples.
- For other relevant terms see:

Glossary of gene expression terms

Protocol (natural sciences)

External links

- Many important links can be found at the Open Directory Project
 - Gene Expression ^[22] at the Open Directory Project
 - Micro Scale Products and Services for Biochemistry and Molecular Biology ^[23] at the Open Directory Project
 - Products and Services for Gene Expression ^[24] at the Open Directory Project
- PLoS Biology Primer: Microarray Analysis ^[25]
- Rundown of microarray technology ^[26]
- ArrayMining.net ^[27] - a free web-server for online microarray analysis
- CLASSIFI ^[28] - Gene Ontology-based gene cluster classification resource
- Microarray - How does it work? ^[29]
- What Are DNA Microarrays ^[30] - A Non-Biologists Introduction to Microarrays
- Microarray data processing using Self-Organizing Maps tutorial: Part 1 ^[31] Part 2 ^[32]
- PNAS Commentary: Discovery of Principles of Nature from Mathematical Modeling of DNA Microarray Data ^[33]

References

- [1] Kulesh DA, Clive DR, Zarlenga DS, Greene JJ (1987). "Identification of interferon-modulated proliferation-related cDNA sequences". *Proc Natl Acad Sci USA* **84** (23): 8453–8457. doi:10.1073/pnas.84.23.8453. PMID 2446323.
- [2] Schena M, Shalon D, Davis RW, Brown PO (1995). "Quantitative monitoring of gene expression patterns with a complementary DNA microarray". *Science* **270** (5235): 467–470. doi:10.1126/science.270.5235.467. PMID 7569999.
- [3] Lashkari DA, DeRisi JL, McCusker JH, Namath AF, Gentile C, Hwang SY, Brown PO, Davis RW (1997). "Yeast microarrays for genome wide parallel genetic and gene expression analysis". *Proc Natl Acad Sci USA* **94** (24): 13057–13062. doi:10.1073/pnas.94.24.13057. PMID 9371799.
- [4] Adomas A, Heller G, Olson A, Osborne J, Karlsson M, Nahalkova J, Van Zyl L, Sederoff R, Stenlid J, Finlay R, Asiegbu FO (2008). "Comparative analysis of transcript abundance in *Pinus sylvestris* after challenge with a saprotrophic, pathogenic or mutualistic fungus". *Tree Physiol.* **28** (6): 885–897. PMID 18381269.
- [5] Pollack JR, Perou CM, Alizadeh AA, Eisen MB, Pergamenschikov A, Williams CF, Jeffrey SS, Botstein D, Brown PO (1999). "Genome-wide analysis of DNA copy-number changes using cDNA microarrays". *Nat Genet* **23** (1): 41–46. doi:10.1038/14385. PMID 10471496.
- [6] Moran G, Stokes C, Thewes S, Hube B, Coleman DC, Sullivan D (2004). "Comparative genomics using *Candida albicans* DNA microarrays reveals absence and divergence of virulence-associated genes in *Candida dubliniensis*". *Microbiology* **150** (Pt 10): 3363–3382. doi:10.1099/mic.0.27221-0. PMID 15470115.
- [7] <http://bgmo.jrc.ec.europa.eu/home/docs.htm>
- [8] Hacia JG, Fan JB, Ryder O, Jin L, Edgemon K, Ghandour G, Mayer RA, Sun B, Hsie L, Robbins CM, Brody LC, Wang D, Lander ES, Lipshutz R, Fodor SP, Collins FS (1999). "Determination of ancestral alleles for human single-nucleotide polymorphisms using high-density oligonucleotide arrays". *Nat Genet* **22**: 164–167. doi:10.1038/9674. PMID 10369258.
- [9] Lausted C et al. (2004). "POSaM: a fast, flexible, open-source, inkjet oligonucleotide synthesizer and microarrayer" (<http://genomebiology.com/2004/5/8/R58>). *Genome Biology* **5** (8): R58. doi:10.1186/gb-2004-5-8-r58. PMID 15287980. .
- [10] Bammler T, Beyer RP; Consortium, Members of the Toxicogenomics Research (2005). "Standardizing global gene expression analysis between laboratories and across platforms". *Nat Methods* **2** (5): 351–356. doi:10.1038/nmeth0605-477a. PMID 15846362.
- [11] Pease AC, Solas D, Sullivan EJ, Cronin MT, Holmes CP, Fodor SP. (1994). "Light-generated oligonucleotide arrays for rapid DNA sequence analysis". *PNAS* **91** (11): 5022–5026. doi:10.1073/pnas.91.11.5022. PMID 8197176.
- [12] Nuwaysir EF, Huang W, Albert TJ, Singh J, Nuwaysir K, Pitas A, Richmond T, Gorski T, Berg JP, Ballin J, McCormick M, Norton J, Pollock T, Sumwalt T, Butcher L, Porter D, Molla M, Hall C, Blattner F, Sussman MR, Wallace RL, Cerrina F, Green RD. (2002). "Gene expression analysis using oligonucleotide arrays produced by maskless photolithography". *Genome Res* **12** (11): 1749–1755. doi:10.1101/gr.362402. PMID 12421762.
- [13] Shalon D, Smith SJ, Brown PO (1996). "A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization". *Genome Res* **6** (7): 639–645. doi:10.1101/gr.6.7.639. PMID 8796352.
- [14] Tang T, François N, Glatigny A, Agier N, Mucchielli MH, Aggerbeck L, Delacroix H (2007). "Expression ratio evaluation in two-colour microarray experiments is significantly improved by correcting image misalignment". *Bioinformatics* **23** (20): 2686–2691. doi:10.1093/bioinformatics/btm399. PMID 17698492.
- [15] Churchill, GA (2002). "Fundamentals of experimental design for cDNA microarrays" (<http://www.vmr.org/research-websites/gcf/Forms/Churchill.pdf>) (– Scholar search (http://scholar.google.co.uk/scholar?hl=en&lr=&q=intitle:Fundamentals+of+experimental+design+for+cDNA+microarrays+as_publication=Nature+genetics+suppliment&as_ylo=2002&as_yhi=2002&btnG=Search)). *Nature genetics supplement* **32**: 490. doi:10.1038/ng1031. PMID 12454643. .
- [16] NCTR Center for Toxicoinformatics - MAQC Project (<http://www.fda.gov/nctr/science/centers/toxicoinformatics/maqc/>)
- [17] Yuk Fai Leung and Duccio Cavalieri, Fundamentals of cDNA microarray data analysis. TRENDS in Genetics Vol.19 No.11 November 2003
- [18] Wei C, Li J, Bumgarner RE. (2004). "Sample size for detecting differentially expressed genes in microarray experiments". *BMC Genomics* **5** (1): 87. doi:10.1186/1471-2164-5-87. PMID 15533245.
- [19] Emmert-Streib, F. and Dehmer, M. (2008). *Analysis of Microarray Data A Network-Based Approach*. Wiley-VCH. ISBN 3-527-31822-4.
- [20] Wouters L, Göhlmann HW, Bijmens L, Kass SU, Molenberghs G, Lewi PJ (2003). "Graphical exploration of gene expression data: a comparative study of three multivariate methods". *Biometrics* **59**: 1131–1139. doi:10.1111/j.0006-341X.2003.00130.x.
- [21] Jain N, Thatte J, Braciale T, Ley K, O'Connell M, Lee JK (2003). "Local-pooled-error test for identifying differentially expressed genes with a small number of replicated microarrays". *Bioinformatics* **19**: 1945–1951. doi:10.1093/bioinformatics/btg264.
- [22] http://www.dmoz.org/Science/Biology/Biochemistry_and_Molecular_Biology/Gene_Expression/
- [23] http://www.dmoz.org/Science/Biology/Biochemistry_and_Molecular_Biology/Products_and_Services/Micro_Scale/
- [24] http://www.dmoz.org/Science/Biology/Biochemistry_and_Molecular_Biology/Gene_Expression/Products_and_Services/
- [25] <http://biology.plosjournals.org/perlserv/?request=get-document&doi=10.1371/journal.pbio.0000015>
- [26] <http://www.genome.gov/page.cfm?pageID=10000533>
- [27] <http://www.arraymining.net>
- [28] <http://pathcuric1.swmed.edu/pathdb/classifi.html>
- [29] http://www.unsolvedmysteries.oregonstate.edu/microarray_07

- [30] <http://www.bioinformaticstutorials.com/?p=8>
- [31] <http://blog.peltarion.com/2007/04/10/the-self-organized-gene-part-1>
- [32] <http://blog.peltarion.com/2007/06/13/the-self-organized-gene-part-2>
- [33] <http://www.pnas.org/content/103/44/16063.extract>

Triple-stranded DNA

A **triple-stranded DNA** is a structure of DNA in which three oligonucleotides wind around each other and form a triple helix. In this structure, one strand binds to a B-form DNA double helix through Hoogsteen or reversed Hoogsteen hydrogen bonds.

For example, a nucleobase T binds to a Watson-Crick base-pairing of T-A by Hoogsteen hydrogen bonds between an AxT pair (x represents a Hoogsteen base pair). An N-3 protonated cytosine, represented as C⁺, can also form a base-triplet with a C-G pair through the Hoogsteen base-pairing of an GxC⁺. Thus, the triple-helical DNAs using these Hoogsteen pairings consist of two homopyrimidines and one homopurine, and the homopyrimidine third strand is parallel to the homopurine strand.

A homopurine third strand can also bind to a homopurine-homopyrimidine duplex using reversed Hoogsteen patterns. In this triplex, a nucleobase A binds to a T-A base pair and a G to a C-G pair. Since the nucleobases on the third strand have to be reversed, the homopurine third strand is antiparallel to the homopurine strand of the original duplex.

Triple-stranded DNA was a common hypothesis in the 1950s when scientists were struggling to discover DNA's true structural form. Watson and Crick (who later won the Nobel Prize for their double-helix model) originally considered a triple-helix model, as did Pauling and Corey who published a proposal for their triple-helix model in the 1953 scientific journal *Nature*, as well as fellow scientist Fraser. However, Watson and Crick soon identified several problems with these models: 1) Negatively charged phosphates near the axis will repel each other, leaving the question as to how the three-chain structure would stay together. 2) In a triple-helix model (specifically Pauling and Corey's model), some of the van der Waals distances appear to be too small. Fraser's model differed from Pauling and Corey's in that in his model the phosphates are on the outside and the bases are on the inside, linked together by hydrogen bonds. However, Watson and Crick found Fraser's model to be too ill-defined to comment specifically on its inadequacies in their publication in "Nature" (1953): *Molecular Structure of Nucleic Acids*.

Triple-stranded DNA was also described in 1957, when it was thought to occur in only one *in vivo* biological process: as an intermediate product during the action of the *E. coli* recombination enzyme RecA. Its role in that process is not understood.

Using nucleic acid segments that bind to the DNA duplexes to form triple strands as a way of regulating gene expression is under investigation by biotechnology companies. Similar work is also being undertaken at Yale University.

References

- Rich, A. (1993). DNA comes in many forms. *Gene* 135:99-109
- Soyfer, V.N and V.N.Potaman. Triple-Helical Nucleic Acids. Springer Ver., New York, 1995, 360 pp.
- Mills, M., Arimondo, P., Lacroix, L., Garestier, T., Hélène, C., Klump, H. & Mergny, J.L. (1999). Energetics of strand displacement reactions in triple helices: a spectroscopic study. ^[1] *J. Mol. Biol.* Vol. 291: 1035-1054
- Watson, J.D. and Crick, F.H.C. (1953). Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. "Nature," 4356: 737-734.

References

- [1] http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6WK7-45R874W-7C&_coverDate=09%2F03%2F1999&_alid=432319574&_rdoc=1&_fmt=&_orig=search&_qd=1&_cdi=6899&_sort=d&view=c&_acct=C000033878&_version=1&_urlVersion=0&_userid=635696&md5=48714a2a903a6741b9ec78e14b98de9f

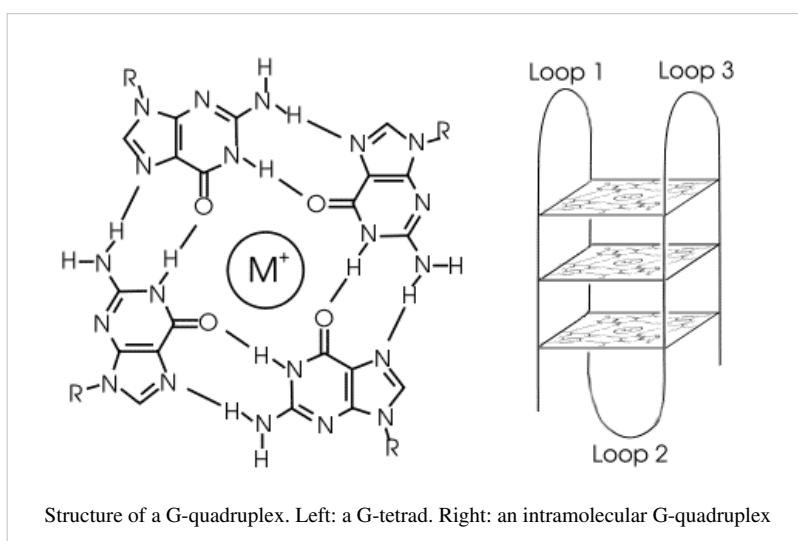
G-quadruplex

Nucleic acid sequences which are rich in guanine are capable of forming four-stranded structures called G-quadruplexes (also known as G-tetrads or G_4 -DNA). These consist of a square arrangement of guanines (a tetrad), stabilized by Hoogsteen hydrogen bonding. They are further stabilized by the existence of a monovalent cation (especially potassium) in the center of the tetrads. They can be formed of DNA, RNA, LNA and PNA, and may be intramolecular, bimolecular or tetramolecular. Depending on the direction of the strands or parts of a strand that form the tetrads, structures may be described as parallel or antiparallel.

Telomeric quadruplexes

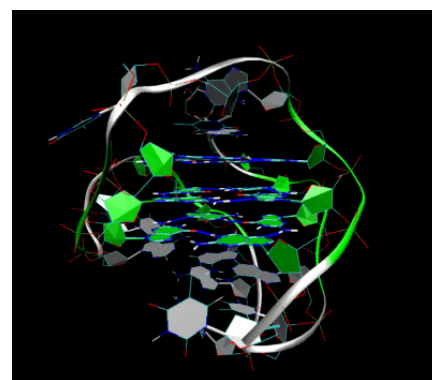
Telomeric repeats in a variety of organisms have been shown to form these structures *in vitro*, and they have also been shown to form *in vivo* in some cases. The human telomeric repeat (which is the same for all vertebrates) consists of many repeats of the sequence d(GGTTAG), and the quadruplexes formed by this structure have been well studied by NMR and X-ray crystal structure determination. The formation of these quadruplexes in telomeres has been shown to decrease

the activity of the enzyme telomerase, which is responsible for maintaining length of telomeres and is involved in around 85% of all cancers. This is an active target of drug discovery.



Non-telomeric quadruplexes

Recently, there has been increasing interest in quadruplexes in locations other than at the telomere. This was given a large boost by the work by Simonsson [1] and Hurley [2] on the proto-oncogene *c-myc*, which was shown to form a quadruplex in a nuclease hypersensitive region critical for gene activity. Since then, many other genes have been shown to have G-quadruplexes in their promoter regions, including the chicken β -globin gene, human ubiquitin-ligase RFP2 and the proto-oncogenes *c-kit*, *bcl-2*, VEGF, H-ras and N-ras. This list is ever-increasing.



3D Structure of the intramolecular human telomeric G-quadruplex in potassium solution (PDB ID 2HY9). The backbone is represented by a tube. The center of this structure contains three layers of G-tetrads. The hydrogen bonds in these layers are represented by blue dashed lines.

Genome-wide surveys based on a quadruplex folding rule have been performed, which have identified 376,000 Putative Quadruplex Sequences (PQS) in the human genome, although not all of these probably form *in vivo*. [3] A similar study has identified putative G-quadruplexes in prokaryotes [4]. There are several possible models for how quadruplexes could control gene activity, either by upregulation or downregulation. One model is shown below, with G-quadruplex formation in or near a promoter blocking transcription of the gene, and hence de-activating it. In another model, quadruplex formed at the non-coding DNA strand helps to maintain an open conformation of the coding DNA strand and enhance an expression of the respective gene.

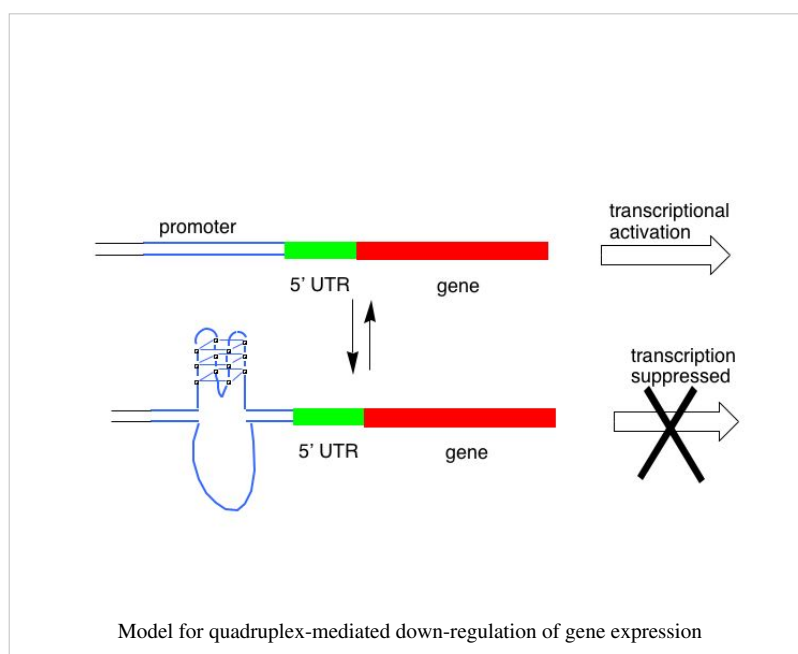
Ligands which bind quadruplexes

One way of inducing or stabilizing G-quadruplex formation, is to introduce a molecule which can bind to the G-quadruplex structure, and a number of ligands, both small molecules and proteins, have been developed which can do so. This has become an increasingly large field of research.

A number of naturally occurring proteins have been identified which selectively bind to G-quadruplexes. These include the helicases implicated in Bloom's and Werner's syndromes

and the *Saccharomyces cerevisiae* protein RAP1. An artificially derived three zinc finger protein called Gq1, which is specific for G-quadruplexes has also been developed, as have specific antibodies.

Cationic porphyrins have been shown to bind intercalatively with G-quadruplexes, as well as the molecule telomestatin.



Quadruplex prediction techniques

Identifying and predicting sequences which have the capacity to form quadruplexes is an important tool in further understanding of their role. A rule for predicting the formation has been proposed, where sequences are predicted to fold based on the pattern $d(G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+})$, where N is any base (including guanine). This rule has been widely used in on-line algorithms.

External links

Books

- Quadruplex Nucleic Acids (ISBN 0-85404-374-8) Neidle & Balasubramanian (Eds.) 2006 ^[5]

Quadruplex websites

- Quadruplex.org ^[6] - a website to serve the quadruplex community
- Quadbase ^[7] - downloadable data on predicted G-quadruplexes
- Greglist ^[8] - a database listing potential G-quadruplex regulated genes
- Database on Quadruplex information: QuadBase from IGIB ^[9]
- GRSDb ^[10] - a database of G-quadruplexes near RNA processing sites.
- GRS_UTRdb ^[11] - a database of G-quadruplexes in the UTRs.
- G-quadruplex Resource Site ^[12]

Research papers

- In vivo veritas: Using yeast to probe the biological functions of G-quadruplexes. Johnson JE, Smith JS, Kozak ML, Johnson FB. **Biochimie**. 2008 Feb 21 ^[13]
- Prevalence of quadruplexes in the human genome, Huppert and Balasubramanian, **NAR** 2005 33(9) 2908-2916 ^[3]
- Highly prevalent putative quadruplex sequence motifs in human DNA, Todd, Johnston, and Neidle, **NAR** 2005 33(9) 2901-2907 ^[14]
- Quadruplex DNA: sequence, topology and structure, Burge, Parkinson, Hazel, Todd and Neidle, **NAR** 2006 34(19) 5402-5415 ^[15]
- Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription, Siddiqui-Jain *et al.*, **PNAS** 2002 99(18) 11593-8 ^[2]
- Genome-wide prediction of G4 DNA as regulatory motifs: Role in Escherichia coli global regulation, Rawal P, Kumarasetti VB, Ravindran J, Kumar N, Halder K, Sharma R, Mukerji M, Das SK, Chowdhury S., **Genome Res.** 2006 16(5) 644-55 ^[4]
- A Biomimetic Potassium Responsive Nanochannel: G-Quadruplex DNA Conformational Switching in a Synthetic Nanopore, Xu Hou, Wei Guo, Fan Xia, Fu-Qiang Nie, Hua Dong, Ye Tian, Liping Wen, Lin Wang, Liuxuan Cao, Yang Yang, Jianming Xue, Yanlin Song, Yugang Wang, Dongsheng Liu and Lei Jiang, **J. Am. Chem. Soc.**, 2009 131(22) 7800–7805 ^[16]

Tools to predict G-quadruplex motifs

- Quadparser: Downloadable program for finding putative quadruplex-forming sequences^[17]
- QGRS Mapper: a web-based application for predicting G-quadruplexes in nucleotide sequences and NCBI genes^[18]
- Quadfinder: Tool for Prediction and Analysis of G Quadruplex Motifs in DNA/RNA Sequences from IGIB^[19]

References

- [1] <http://nar.oxfordjournals.org/cgi/content/full/26/5/1167?maxtoshow=&HITS=10&hits=10&RESULTFORMAT=1&author1=simonsson%2C+t&andorexacttitle=and&andorexacttitleabs=and&andorexactfulltext=and&searchid=1&FIRSTINDEX=0&sortspec=relevance&resourcetype=HWCIT>
- [2] <http://www.pnas.org/cgi/content/abstract/99/18/11593>
- [3] <http://nar.oxfordjournals.org/cgi/content/abstract/33/9/2908>
- [4] <http://www.genome.org/cgi/content/short/16/5/644>
- [5] <http://chemistry.rsc.org/Publishing/Books/0854043748.asp>
- [6] <http://www.quadruplex.org>
- [7] <http://www.quadruplex.org/?view=quadbase>
- [8] <http://tubic.tju.edu.cn/greglist/>
- [9] <http://quadbase.igib.res.in>
- [10] <http://bioinformatics.ramapo.edu/GRSDB2/>
- [11] http://bioinformatics.ramapo.edu/GRS_UTRDB/
- [12] <http://bioinformatics.ramapo.edu/QGRS/>
- [13] [http://linkinghub.elsevier.com/retrieve/pii/S0300-9084\(08\)00047-3](http://linkinghub.elsevier.com/retrieve/pii/S0300-9084(08)00047-3)
- [14] <http://nar.oxfordjournals.org/cgi/content/abstract/33/9/2901>
- [15] <http://nar.oxfordjournals.org/cgi/content/abstract/34/19/5402>
- [16] <http://pubs.acs.org/doi/abs/10.1021/ja901574c>
- [17] <http://www-shankar.ch.cam.ac.uk/quadparser.html>
- [18] <http://bioinformatics.ramapo.edu/QGRS/>
- [19] <http://202.54.26.221/quadfinder/>

DNA Analysis

Genetic Testing : Gene tests (also called DNA-based tests), the newest and most sophisticated of the techniques used to test for genetic disorders, involve direct examination of the DNA molecule itself. Other genetic tests include biochemical tests for such gene products as enzymes and other proteins and for microscopic examination of stained or fluorescent chromosomes. Genetic tests are used for several reasons, including:

- carrier screening, which involves identifying unaffected individuals who carry one copy of a gene for a disease that requires two copies for the disease to be expressed
- preimplantation genetic diagnosis (see the side bar, Screening Embryos for Disease)
- prenatal diagnostic testing
- newborn screening
- presymptomatic testing for predicting adult-onset disorders such as Huntington's disease
- presymptomatic testing for estimating the risk of developing adult-onset cancers and Alzheimer's disease
- confirmational diagnosis of a symptomatic individual
- forensic/identity testing

Genetic testing allows the genetic diagnosis of vulnerabilities to inherit diseases, and can also be used to determine a child's paternity (genetic father) or a person's ancestry. Normally, every person carries two copies of every gene, one inherited from their mother, one inherited from their father. The human genome is believed to contain around 20,000 - 25,000 genes. In addition to studying chromosomes to the level of individual genes, genetic testing in a broader sense includes biochemical tests for the possible presence of genetic diseases, or mutant forms of genes associated with increased risk of developing genetic disorders. Genetic testing identifies changes in chromosomes, genes, or proteins.^[1] Most of the time, testing is used to find changes that are associated with inherited disorders. The results of a genetic test can confirm or rule out a suspected genetic condition or help determine a person's chance of developing or passing on a genetic disorder. Several hundred genetic tests are currently in use, and more are being developed.^{[2] [3]}

Since genetic testing may open up ethical or psychological problems, genetic testing is often accompanied by genetic counseling.

Types

Genetic testing is "the analysis of, chromosomes (DNA), proteins, and certain metabolites in order to detect heritable disease-related genotypes, mutations, phenotypes, or karyotypes for clinical purposes."^[4] It can provide information about a person's genes and chromosomes throughout life. Available types of testing include:

- **Newborn screening:** Newborn screening is used just after birth to identify genetic disorders that can be treated early in life. The routine testing of infants for certain disorders is the most widespread use of genetic testing—millions of babies are tested each year in the United States. All states currently test infants for phenylketonuria (a genetic disorder that causes mental illness if left untreated) and congenital hypothyroidism (a disorder of the thyroid gland).
- **Diagnostic testing:** Diagnostic testing is used to diagnose or rule out a specific genetic or chromosomal condition. In many cases, genetic testing is used to confirm a diagnosis when a particular condition is suspected based on physical mutations and symptoms. Diagnostic testing can be performed at any time during a person's life, but is not available for all genes or all genetic conditions. The results of a diagnostic test can influence a person's choices about health care and the management of the disease.
- **Carrier testing:** Carrier testing is used to identify people who carry one copy of a gene mutation that, when present in two copies, causes a genetic disorder. This type of testing is offered to individuals who have a family history of a genetic disorder and to people in ethnic groups with an increased risk of specific genetic conditions. If both

parents are tested, the test can provide information about a couple's risk of having a child with a genetic condition.

- **Prenatal testing:** Prenatal testing is used to detect changes in a fetus's genes or chromosomes before birth. This type of testing is offered to couples with an increased risk of having a baby with a genetic or chromosomal disorder. In some cases, prenatal testing can lessen a couple's uncertainty or help them decide whether to abort the pregnancy. It cannot identify all possible inherited disorders and birth defects, however.
- **Preimplantation genetic diagnosis:** Genetic testing procedures that are performed on human embryos prior to the implantation as part of an in vitro fertilization procedure.
- **Predictive and presymptomatic testing:** Predictive and presymptomatic types of testing are used to detect gene mutations associated with disorders that appear after birth, often later in life. These tests can be helpful to people who have a family member with a genetic disorder, but who have no features of the disorder themselves at the time of testing. Predictive testing can identify mutations that increase a person's chances of developing disorders with a genetic basis, such as certain types of cancer. For example, an individual with a mutation in *BRCA1* has a 65% cumulative risk of breast cancer [5]. Presymptomatic testing can determine whether a person will develop a genetic disorder, such as hemochromatosis (an iron overload disorder), before any signs or symptoms appear. The results of predictive and presymptomatic testing can provide information about a person's risk of developing a specific disorder and help with making decisions about medical care.
- **Forensic testing:** Forensic testing uses DNA sequences to identify an individual for legal purposes. Unlike the tests described above, forensic testing is not used to detect gene mutations associated with disease. This type of testing can identify crime or catastrophe victims, rule out or implicate a crime suspect, or establish biological relationships between people (for example, paternity).
- **Parental testing:** This type of genetic test uses special DNA markers to identify the same or similar inheritance patterns between related individuals. Based on the fact that we all inherit half of our DNA from the father, and half from the mother, DNA scientists test individuals to find the match of DNA sequences at some highly differential markers to draw the conclusion of relatedness.
- **Research testing:** Research testing includes finding unknown genes, learning how genes work and advancing our understanding of genetic conditions. The results of testing done as part of a research study are usually not available to patients or their healthcare providers.

Medical procedure

Genetic testing is often done as part of a genetic consultation and as of mid-2008 there were more than 1,200 clinically applicable genetic tests available.^[6] Once a person decides to proceed with genetic testing, a medical geneticist, genetic counselor, primary care doctor, or specialist can order the test after obtaining informed consent.

Genetic tests are performed on a sample of blood, hair, skin, amniotic fluid (the fluid that surrounds a fetus during pregnancy), or other tissue. For example, a medical procedure called a buccal smear uses a small brush or cotton swab to collect a sample of cells from the inside surface of the cheek. Alternatively, a small amount of saline mouthwash may be swished in the mouth to collect the cells. The sample is sent to a laboratory where technicians look for specific changes in chromosomes, DNA, or proteins, depending on the suspected disorder. The laboratory reports the test results in writing to a person's doctor or genetic counselor.

Routine newborn screening tests are done on a small blood sample obtained by pricking the baby's heel with a lancet.

Interpreting results

The results of genetic tests are not always straightforward, which often makes them challenging to interpret and explain. When interpreting test results, healthcare professionals consider a person's medical history, family history, and the type of genetic test that was done.

A positive test result means that the laboratory found a change in a particular gene, chromosome, or protein of interest. Depending on the purpose of the test, this result may confirm a diagnosis, indicate that a person is a carrier of a particular genetic mutation, identify an increased risk of developing a disease (such as cancer) in the future, or suggest a need for further testing. Because family members have some genetic material in common, a positive test result may also have implications for certain blood relatives of the person undergoing testing. It is important to note that a positive result of a predictive or presymptomatic genetic test usually cannot establish the exact risk of developing a disorder. Also, health professionals typically cannot use a positive test result to predict the course or severity of a condition.

A negative test result means that the laboratory did not find a dangerous copy of the gene, chromosome, or protein under consideration. This result can indicate that a person is not affected by a particular disorder, is not a carrier of a specific genetic mutation, or does not have an increased risk of developing a certain disease. It is possible, however, that the test missed a disease-causing genetic alteration because many tests cannot detect all genetic changes that can cause a particular disorder. Further testing may be required to confirm a negative result.

In some cases, a negative result might not give any useful information. This type of result is called uninformative, indeterminate, inconclusive, or ambiguous. Uninformative test results sometimes occur because everyone has common, natural variations in their DNA, called polymorphisms, that do not affect health. If a genetic test finds a change in DNA that has not been associated with a disorder in other people, it can be difficult to tell whether it is a natural polymorphism or a disease-causing mutation. An uninformative result cannot confirm or rule out a specific diagnosis, and it cannot indicate whether a person has an increased risk of developing a disorder. In some cases, testing other affected and unaffected family members can help clarify this type of result.

Risks and limitations

The physical risks associated with most genetic tests are very small, particularly for those tests that require only a blood sample or buccal smear (a procedure that samples cells from the inside surface of the cheek). The procedures used for prenatal testing carry a small but real risk of losing the pregnancy (miscarriage) because they require a sample of amniotic fluid or tissue from around the fetus.

Many of the risks associated with genetic testing involve the emotional, social, or financial consequences of the test results. People may feel angry, depressed, anxious, or guilty about their results. In some cases, genetic testing creates tension within a family because the results can reveal information about other family members in addition to the person who is tested. The possibility of genetic discrimination in employment or insurance is also a concern. Some individuals avoid genetic testing out of fear it will affect their ability to purchase insurance or find a job.^[7] Health insurers do not currently require applicants for coverage to undergo genetic testing, and when insurers encounter genetic information, it is subject to the same confidentiality protections as any other sensitive health information.^[8] In the United States, the use of genetic information is governed by the Genetic Information Nondiscrimination Act (GINA) (see discussion below in "U.S. Government Regulation Related to Genetic Testing and Information").

Genetic testing can provide only limited information about an inherited condition. The test often can't determine if a person will show symptoms of a disorder, how severe the symptoms will be, or whether the disorder will progress over time. Another major limitation is the lack of treatment strategies for many genetic disorders once they are diagnosed.

A genetics professional can explain in detail the benefits, risks, and limitations of a particular test. It is important that any person who is considering genetic testing understand and weigh these factors before making a decision.

Direct-to-Consumer (DTC) Genetic Testing

Direct-to-Consumer (DTC) genetic testing is a type of genetic test that is accessible directly to the consumer without having to go through a health care professional. Usually, to obtain a genetic test, health care professionals such as doctors acquire the permission of the patient and order the desired test. DTC genetic tests, however, allow consumers to bypass this process and order one themselves. There are a variety of DTC tests, ranging from testing for breast cancer alleles to mutations linked to cystic fibrosis. Benefits of DTC testing are the accessibility of tests to consumers, promotion of proactive healthcare and the privacy of genetic information. Possible additional risks of DTC testing are the lack of governmental regulation and the potential misinterpretation of genetic information.

Controversy

DTC genetic testing has been controversial due to outspoken opposition within the scientific community. Critics of DTC testing argue against the risks involved, the unregulated advertising and marketing claims, and the overall lack of governmental oversight.^[9]

DTC testing involves many of the same risks associated with any genetic test. One of the more obvious and dangerous of these is the possibility of severe misreading of test results. Without professional guidance, consumers can potentially misinterpret genetic information, causing them to be deluded about their personal health.

Some advertising for direct-to-consumer genetic testing has been criticized as conveying an exaggerated and inaccurate message about the connection between genetic information and disease risk, utilizing emotions as a selling factor. An advertisement for a BRCA-predictive genetic test for breast cancer stated: "There is no stronger antidote for fear than information."^[10]

U.S. Government Regulation Related to Genetic Testing and Information

Currently, the U.S. has no strong Federal regulation moderating the DTC market. Though there are several hundred tests available, only a handful are approved by the Food and Drug Administration (FDA); these are sold as at-home test kits, and are therefore considered "medical devices" over which the FDA may assert jurisdiction. Other types of DTC tests require customers to mail in DNA samples for testing; it is difficult for the FDA to exercise jurisdiction over these types of tests, because the actual testing is completed in the laboratories of providers. As of 2007, the FDA had not yet officially substantiated with scientific evidence the claimed accuracy of the majority of direct-to-consumer genetic tests.^[11]

With regard to genetic testing and information in general, legislation in the United States called the Genetic Information Nondiscrimination Act prohibits group health plans and health insurers from denying coverage to a healthy individual or charging that person higher premiums based solely on a genetic predisposition to developing a disease in the future. The legislation also bars employers from using individuals' genetic information when making hiring, firing, job placement, or promotion decisions.^[12] The legislation, the first of its kind in the U.S.,^[13] was passed by the United States Senate on April 24, 2008, on a vote of 95-0, and was signed into law by President George W. Bush on May 21, 2008.^[14] ^[15] It went into effect on November 21, 2009.

Fiction

Some possible future ethical problems of genetic testing were considered in the science fiction film *Gattaca*, and the science fiction anime series "Gundam Seed". Also some films which include the topic of genetic testing include, "The Island" and the "Resident Evil" series.

See also

- Full Genome Sequencing
- List of human genes
- List of genetic disorders
- Gene theft

External links

- GeneTests^[16] US National Institutes of Health funded resource on genetic testing.
- EuroGentest^[17] European network for test development, harmonization, validation and standardization
- TECHGENE^[18] European project on genetic testing and next generation sequencing technology.
- "The science and practice of genetic testing for Huntington's disease"^[19]. Huntington's Disease Outreach Project for Education at Stanford
- Downloadable article: "Evidence that a West-East admixed population lived in the Tarim Basin as early as the early Bronze Age" Li et al. *BMC Biology* 2010, 8:15. [20]
- KnowYourGenes.org^[21] Genetic Disease Foundation

References

- [1] http://www.ghr.nlm.nih.gov/handbook/testing/genetic_testing
- [2] http://www.nlm.nih.gov/medlineplus/genetic_testing.html
- [3] "Definitions of Genetic Testing" (http://www.eurogentest.org/patient/public_health/info/public/unit3/DefinitionsGeneticTesting-3rdDraf18Jan07.xhtml). *Definitions of Genetic Testing* (Jorge Sequeiros and Bárbara Guimarães). EuroGentest Network of Excellence Project. 2008-09-11. . Retrieved 2008-08-10.
- [4] Holtzman NA, Murphy PD, Watson MS, Barr PA (October 1997). "Predictive genetic testing: from basic research to clinical practice" (<http://www.sciencemag.org/cgi/pmidlookup?view=long&pmid=9381169>). *Science (journal)* **278** (5338): 602–5. PMID 9381169. .
- [5] http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=12677558
- [6] Allingham-Hawkins, Diane (2008-08-01). "Successful Genetic Tests Are Predicated on Clinical Utility" (<http://www.genengnews.com/articles/chitem.aspx?aid=2544>). *Genetic Engineering & Biotechnology News* (Mary Ann Liebert) **28** (14): pp. 6, 9. ISSN 1935-472X. . Retrieved 2008-09-23.
- [7] Amy Harmon, "Insurance Fears Lead Many to Shun DNA Tests," (http://www.nytimes.com/2008/02/24/health/24dna.html?_r=2&scp=5&sq=Genetic&st=nyt&oref=slogin&oref=slogin) The New York Times, February 24, 2008
- [8] "Genetic Information and Medical Expense Insurance" (<http://www.actuary.org/pdf/health/geneticmono.pdf>), American Academy of Actuaries, June 2000
- [9] Hunter et al., "Letting the Genome out of the Bottle" (<http://content.nejm.org/cgi/content/full/358/2/105>) New England Journal of Medicine
- [10] Gollust et al., "Limitations of Direct-to-Consumer Advertising for Clinical Genetic Testing," (<http://jama.ama-assn.org/cgi/content/full/288/14/1762#REF-JSC20258-14>) JAMA.2002; 288: 1762-1767
- [11] Shawna Williams and Gail Javitt, "Direct-to-consumer genetic testing: empowering or endangering the public?," (http://www.dnapolicy.org/policy.issue.php?action=detail&issuebrief_id=32) The Genetics and Public Policy Center, July 25, 2006 (updated 6/15/2007)
- [12] Statement of Administration policy (<http://www.genome.gov/Pages/PolicyEthics/GeneticDiscrimination/SAPonHR493.pdf>), Executive Office of the President, Office of Management and Budget, 27 April 2007
- [13] Kennedy in support of genetic information nondiscrimination bill (http://kenedy.senate.gov/newsroom/press_release.cfm?id=4FCF8E86-4706-4E74-B451-36253C5A425D), April 24, 2008. Last access: 28/05/2008.
- [14] Keim, Brandon (May 21, 2008). "Genetic Discrimination by Insurers, Employers Becomes a Crime" (<http://blog.wired.com/wiredscience/2008/05/the-genetic-inf.html>). *Wired.com*. . Retrieved 2008-05-28.
- [15] "Administration News | President Bush Signs Genetic Nondiscrimination Legislation Into Law," (http://www.kaisernetwork.org/daily_reports/rep_index.cfm?DR_ID=52305) Kaiser Daily Health Policy Report, Kaiser Family Foundation, May 22, 2008

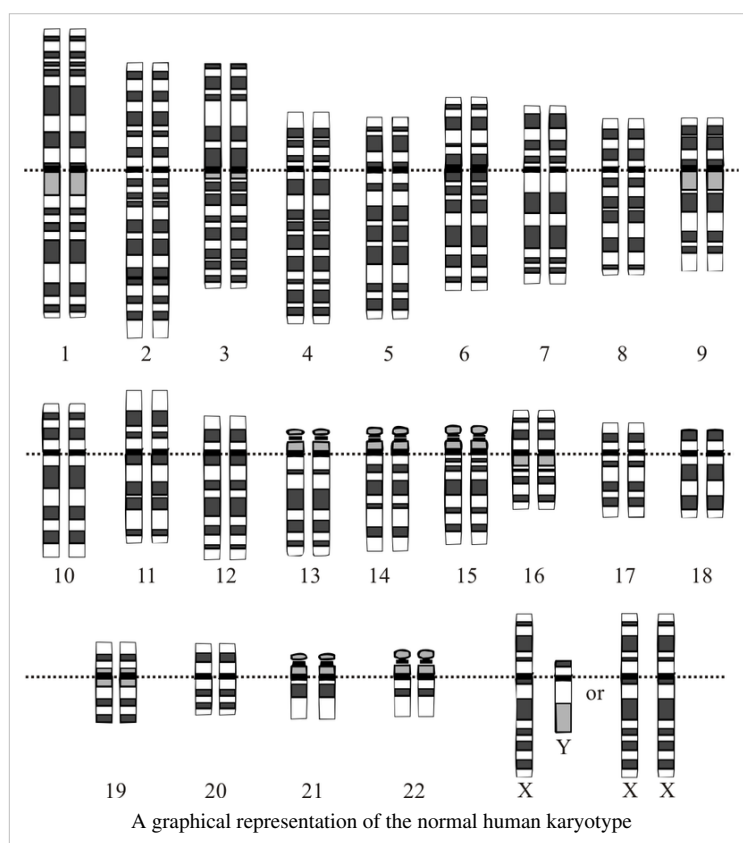
- [16] <http://www.genetests.org>
- [17] <http://www.eurogentest.org>
- [18] <http://www.techgene.eu>
- [19] <http://hopes.stanford.edu/diagnosis/gentest/s0.html>
- [20] <http://www.biomedcentral.com/content/pdf/1741-7007-8-15.pdf>
- [21] <http://www.knowyourgenes.org>

Human genome

The **human genome** is the genome of *Homo sapiens*, which is stored on 23 chromosome pairs. Twenty-two of these are autosomal chromosome pairs, while the remaining pair is sex-determining. The haploid human genome occupies a total of just over 3 billion DNA base pairs. The Human Genome Project (HGP) produced a reference sequence of the euchromatic human genome, which is used worldwide in biomedical sciences.

The haploid human genome contains ca. 23,000 protein-coding genes, far fewer than had been expected before its sequencing.^[1]

^[2] In fact, only about 1.5% of the genome codes for proteins, while the rest consists of non-coding RNA genes, regulatory sequences, introns, and (controversially named) "junk" DNA.^[3]



Features

Genes

There are estimated to be between 20,000 and 25,000 human protein-coding genes. The estimate of the number of human genes has been repeatedly revised down from initial predictions of 100,000 or more as genome sequence quality and gene finding methods have improved.

Surprisingly, the number of human genes seems to be less than a factor of two greater than that of many much simpler organisms, such as the roundworm and the fruit fly. However, human cells make extensive use of alternative splicing to produce several different proteins from a single gene, and the human proteome is thought to be much larger than those of the aforementioned organisms. Besides, most human genes have multiple exons, and human introns are frequently much longer than the flanking exons.

Human genes are distributed unevenly across the chromosomes. Each chromosome contains various gene-rich and gene-poor regions, which seem to be correlated with chromosome bands and GC-content. The significance of these nonrandom patterns of gene density is not well understood. In addition to protein coding genes, the human genome contains thousands of RNA genes, including tRNA, ribosomal RNA, microRNA, and other non-coding RNA genes.

Regulatory sequences

The human genome has many different regulatory sequences which are crucial to controlling gene expression. These are typically short sequences that appear near or within genes. A systematic understanding of these regulatory sequences and how they together act as a gene regulatory network is only beginning to emerge from computational, high-throughput expression and comparative genomics studies. Some types of non-coding DNA are genetic "switches" that do not encode proteins, but do regulate when and where genes are expressed.^[4]

Identification of regulatory sequences relies in part on evolutionary conservation. The evolutionary branch between the primates and mouse, for example, occurred 70–90 million years ago.^[5] So computer comparisons of gene sequences that identify conserved non-coding sequences will be an indication of their importance in duties such as gene regulation.^[6]

Another comparative genomic approach to locating regulatory sequences in humans is the gene sequencing of the puffer fish. These vertebrates have essentially the same genes and regulatory gene sequences as humans, but with only one-eighth the "junk" DNA. The compact DNA sequence of the puffer fish makes it much easier to locate the regulatory genes.^[7]

Other DNA

Protein-coding sequences (specifically, coding exons) comprise less than 1.5% of the human genome.^[3] Aside from genes and known regulatory sequences, the human genome contains vast regions of DNA the function of which, if any, remains unknown. These regions in fact comprise the vast majority, by some estimates 97%, of the human genome size. Much of this is composed of:

Repeat elements

- Tandem repeats
 - Satellite DNA
 - Minisatellite
 - Microsatellite
- Interspersed repeats
 - SINEs
 - LINEs

Transposons

- Retrotransposons
 - LTR
 - Ty1-copia
 - Ty3-gypsy
 - Non-LTR
 - SINEs
 - LINEs
- DNA Transposons

Junk DNA

However, there is also a large amount of sequence that does not fall under any known classification. Much of this sequence may be an evolutionary artifact that serves no present-day purpose, and these regions are sometimes collectively referred to as "junk" DNA. There are, however, a variety of emerging indications that many sequences within are likely to function in ways that are not fully understood. Recent experiments using microarrays have revealed that a substantial fraction of non-genic DNA is in fact transcribed into RNA,^[8] which leads to the possibility that the resulting transcripts may have some unknown function. Also, the evolutionary conservation across the mammalian genomes of much more sequence than can be explained by protein-coding regions indicates that many, and perhaps most, functional elements in the genome remain unknown.^[9] The investigation of the vast quantity of sequence information in the human genome whose function remains unknown is currently a major avenue of scientific inquiry.^[10]

Information content

The 3 billion base pairs of the haploid human genome correspond to an information content of about 750 megabytes, since every base pair can be coded by 2 bits. The entropy rate of the genome differs significantly between coding and non-coding sequences. It is close to the maximum of 2 bits per base pair for the coding sequences (about 45 million base pairs), and between 1.5 and 1.9 bits per base pair for each individual chromosome, except for the Y chromosome, which has an entropy rate below 0.9 bits per base pair.^[11]

Information content of the haploid human genome by chromosome:

	total (XX)	total (XY)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	Y
million base pairs	3,080	3,022	247	243	199	191	181	171	159	146	140	135	134	132	114	106	100	89	79	76	63	62	47	50	155	58
megabytes (raw data)	770	756	61.8	60.7	49.9	47.8	45.2	42.7	39.7	36.6	35.1	33.9	33.6	33.1	28.5	26.6	25.1	22.2	19.7	19.0	16.0	15.6	11.7	12.4	38.7	14.4
megabytes in zipped Human Genome Project file ^[12]	827	819	65.1	68.2	57.4	52.3	51.3	48.8	45.3	38.6	33.9	39.1	39.8	38.8	28.8	26.5	22.9	22.5	22.7	22.2	16.4	18.9	10.4	10.4	38.6	8.0
entropy rate in bits per base pair (Liu, Venkatesh and Maley 2008)	1.70	1.71	1.82	1.80	1.82	1.82	1.83	1.82	1.81	1.83	1.59	1.83	1.84	1.59	1.56	1.53	1.66	1.82	1.87	1.58	1.86	1.82	1.62	1.83	1.80	0.84

Sequencing

DNA sequencing determines the order of the nucleotide bases in a genome.

Composite

The Human Genome Project and a parallel project by Celera Genomics each produced and published a haploid human genome sequence, both of which were a composite of the DNA sequence of several individuals.^[13]

Personal

A personal genome sequence is a complete sequencing of the chemical base pairs that make up the DNA of a single person. Because medical treatments have different effects on different people because of genetic variations such as single-nucleotide polymorphisms (SNPs), the analysis of personal genomes may lead to personalized medical treatment based on individual genotypes.

The completion of the fifth such map was announced in December 2008. The genome mapped was that of a Korean researcher Seong-Jin Kim. Genome maps had previously been completed for Craig Venter of the U.S. in 2007, James Watson of the U.S. in April 2008, and Yang Huanming of China in November 2008 and Dan Stoicescu in January 2008.^{[14] [15] [16]}

Personal genomes had not been sequenced in the Human Genome Project to protect the identity of volunteers who provided DNA samples. That sequence was derived from the DNA of several volunteers from a diverse population.^[17] Another distinction is that the HGP sequence is haploid, however, the sequence maps for Venter and Watson for example are diploid, representing both sets of chromosomes.

Kim's genome had 1.58 million SNPs that had never been reported before and indicates that six out of 10,000 DNA bases are unique to Koreans. Kim's sequence map can be used to assist in building a standard Korean genome, which can then be used to compare the genomes of other Korean individuals for personalized medical treatments.

Mapping

Whereas a genome sequence lists the order of every DNA base in a genome, a genome map identifies the landmarks. A genome map is less detailed than a genome sequence and aids in navigating around the genome.^{[18] [19]}

Variation

An example of a variation map is the HapMap being developed by the International HapMap Project. The HapMap is a haplotype map of the human genome, "which will describe the common patterns of human DNA sequence variation."^[20] It catalogs the patterns of small-scale variations in the genome that involve single DNA letters, or bases.

Researchers published the first sequence-based map of large-scale structural variation across the human genome in the journal *Nature* in May 2008.^{[21] [22]} Large-scale structural variations are differences in the genome among people that range from a few thousand to a few million DNA bases; some are gains or losses of stretches of genome sequence and others appear as re-arrangements of stretches of sequence. These variations include differences in the number of copies individuals have of a particular gene, deletions, translocations and inversions.

Variation

Most studies of human genetic variation have focused on single-nucleotide polymorphisms (SNPs), which are substitutions in individual bases along a chromosome. Most analyses estimate that SNPs occur on average somewhere between every 1 in 100 and 1 in 300 base pairs in the euchromatic human genome, although they do not occur at a uniform density. Thus follows the popular statement that "we are all, regardless of race, genetically 99.9% the same",^[23] although this would be somewhat qualified by most geneticists. For example, a much larger fraction of the genome is now thought to be involved in copy number variation.^[24] A large-scale collaborative effort to catalog SNP variations in the human genome is being undertaken by the International HapMap Project.

The genomic loci and length of certain types of small repetitive sequences are highly variable from person to person, which is the basis of DNA fingerprinting and DNA paternity testing technologies. The heterochromatic portions of the human genome, which total several hundred million base pairs, are also thought to be quite variable within the human population (they are so repetitive and so long that they cannot be accurately sequenced with current technology). These regions contain few genes, and it is unclear whether any significant phenotypic effect results from typical variation in repeats or heterochromatin.

Most gross genomic mutations in gamete germ cells probably result in inviable embryos; however, a number of human diseases are related to large-scale genomic abnormalities. Down syndrome, Turner Syndrome, and a number of other diseases result from nondisjunction of entire chromosomes. Cancer cells frequently have aneuploidy of chromosomes and chromosome arms, although a cause and effect relationship between aneuploidy and cancer has not been established.

Genetic disorders

Most aspects of human biology involve both genetic (inherited) and non-genetic (environmental) factors. Some inherited variation influences aspects of our biology that are not medical in nature (height, eye color, ability to taste or smell certain compounds, etc). Moreover, some genetic disorders only cause disease in combination with the appropriate environmental factors (such as diet). With these caveats, genetic disorders may be described as clinically defined diseases caused by genomic DNA sequence variation. In the most straightforward cases, the disorder can be associated with variation in a single gene. For example, cystic fibrosis is caused by mutations in the CFTR gene, and is the most common recessive disorder in caucasian populations with over 1,300 different mutations known. Disease-causing mutations in specific genes are usually severe in terms of gene function, and are fortunately rare, thus genetic disorders are similarly individually rare. However, since there are many genes that can vary to cause genetic disorders, in aggregate they comprise a significant component of known medical conditions, especially in pediatric medicine. Molecularly characterized genetic disorders are those for which the underlying causal gene has been identified, currently there are approximately 2,200 such disorders annotated in the OMIM database.^[25]

Studies of genetic disorders are often performed by means of family-based studies. In some instances population based approaches are employed, particularly in the case of so-called founder populations such as those in Finland, French-Canada, Utah, Sardinia, etc. Diagnosis and treatment of genetic disorders are usually performed by a geneticist-physician trained in clinical/medical genetics. The results of the Human Genome Project are likely to provide increased availability of genetic testing for gene-related disorders, and eventually improved treatment. Parents can be screened for hereditary conditions and counselled on the consequences, the probability it will be inherited, and how to avoid or ameliorate it in their offspring.

As noted above, there are many different kinds of DNA sequence variation, ranging from complete extra or missing chromosomes down to single nucleotide changes. It is generally presumed that much naturally occurring genetic variation in human populations is phenotypically neutral, i.e. has little or no detectable effect on the physiology of the individual (although there may be fractional differences in fitness defined over evolutionary time frames). Genetic disorders can be caused by any or all known types of sequence variation. To molecularly characterize a new

genetic disorder, it is necessary to establish a causal link between a particular genomic sequence variant and the clinical disease under investigation. Such studies constitute the realm of human molecular genetics.

With the advent of the Human Genome and International HapMap Project, it has become feasible to explore subtle genetic influences on many common disease conditions such as diabetes, asthma, migraine, schizophrenia, etc. Although some causal links have been made between genomic sequence variants in particular genes and some of these diseases, often with much publicity in the general media, these are usually not considered to be genetic disorders *per se* as their causes are complex, involving many different genetic and environmental factors. Thus there may be disagreement in particular cases whether a specific medical condition should be termed a genetic disorder.

Evolution

Comparative genomics studies of mammalian genomes suggest that approximately 5% of the human genome has been conserved by evolution since the divergence of those species approximately 200 million years ago, containing the vast majority of genes.^[10] ^[9] Intriguingly, since genes and known regulatory sequences probably comprise less than 2% of the genome, this suggests that there may be more unknown functional sequence than known functional sequence. A smaller, yet large, fraction of human genes seem to be shared among most known vertebrates. The chimpanzee genome is variously reported to be 94%-98.5% identical to the human genome in scientific literature.^[26] ^[27] On average, a typical human protein-coding gene differs from its chimpanzee ortholog by only two amino acid substitutions; nearly one third of human genes have exactly the same protein translation as their chimpanzee orthologs. A major difference between the two genomes is human chromosome 2, which is equivalent to a fusion product of chimpanzee chromosomes 12 and 13^[28] (later renamed to chromosomes 2A and 2B, respectively).

Humans have undergone an extraordinary loss of olfactory receptor genes during our recent evolution, which explains our relatively crude sense of smell compared to most other mammals. Evolutionary evidence suggests that the emergence of color vision in humans and several other primate species has diminished the need for the sense of smell.^[29]

Mitochondrial genome

The human mitochondrial genome, while usually not included when referring to the "human genome", is of tremendous interest to geneticists, since it undoubtedly plays a role in mitochondrial disease. It also sheds light on human evolution; for example, analysis of variation in the human mitochondrial genome has led to the postulation of a recent common ancestor for all humans on the maternal line of descent. (see Mitochondrial Eve)

Due to the lack of a system for checking for copying errors, Mitochondrial DNA (mtDNA) has a more rapid rate of variation than nuclear DNA. This 20-fold increase in the mutation rate allows mtDNA to be used for more accurate tracing of maternal ancestry. Studies of mtDNA in populations have allowed ancient migration paths to be traced, such as the migration of Native Americans from Siberia or Polynesians from southeastern Asia. It has also been used to show that there is no trace of Neanderthal DNA in the European gene mixture inherited through purely maternal lineage.^[30]

Epigenome

Epigenetics are a variety of features of the human genome that transcend its primary DNA sequence, such as chromatin packaging, histone modifications and DNA methylation, and which are important in regulating gene expression, genome replication and other cellular processes. Basically they are marks on DNA that are influenced within an individual's own life. Over-eating for years will change a person's epigenetics in such a way that children, grandchildren, and the next few generations will be predisposed to a shorter and unhealthier life. Drugs are being researched that can change one's epigenetics, and which when fully developed will be able to "shut off" certain diseases and disorders. Overall epigenetics strengthen and weaken certain genes but are not an actual part of the

DNA.^{[31] [32]}

See also

- Craig Venter#Individual human genome sequenced
- Eugenics
- Eukaryotic chromosome fine structure
- Genetic distance
- Genomic organization
- Human genetic engineering
- Noncoding DNA
- The Genographic Project
- Y-chromosomal Adam

References

- [1] Evolutionary Trajectories of Primate Genes Involved in HIV Pathogenesis, 2 September 2009 (<http://mbe.oxfordjournals.org/cgi/content/abstract/msp197v1>)
- [2] International Human Genome Sequencing Consortium (2004). "Finishing the euchromatic sequence of the human genome." *Nature* **431** (7011): 931–45. doi:10.1038/nature03001. PMID 15496913. (<http://www.nature.com/nature/journal/v431/n7011/full/nature03001.html>)
- [3] International Human Genome Sequencing Consortium (2001). "Initial sequencing and analysis of the human genome." *Nature* **409** (6822): 860–921. doi:10.1038/35057062. PMID 11237011. (<http://www.nature.com/nature/journal/v409/n6822/full/409860a0.html>)
- [4] Carroll, Sean B. et al. (May 2008). "Regulating Evolution", *Scientific American*, pp. 60–67.
- [5] Nei M, Xu P, Glazko G (2001). "Estimation of divergence times from multiprotein sequences for a few mammalian species and several distantly related organisms." (<http://www.pnas.org/cgi/content/full/051611498>). *Proc Natl Acad Sci USA* **98** (5): 2497–502. doi:10.1073/pnas.051611498. PMID 11226267. .
- [6] Loots G, Locksley R, Blankespoor C, Wang Z, Miller W, Rubin E, Frazer K (2000). "Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons." *Science* **288** (5463): 136–40. doi:10.1126/science.288.5463.136. PMID 10753117. Summary (<http://www.lbl.gov/Science-Articles/Archive/mouse-dna-model.html>)
- [7] Meunier, Monique. "[[Genoscope (http://www.cns.fr/externe/English/Actualites/Presse/261001_1.html)] and Whitehead announce a high sequence coverage of the Tetraodon nigroviridis genome"]. Genoscope. . Retrieved 2006-09-12.
- [8] "...a tiling array with 5-nucleotide resolution that mapped transcription activity along 10 human chromosomes revealed that an average of 10% of the genome (compared to the 1 to 2% represented by bona fide exons) corresponds to polyadenylated transcripts, of which more than half do not overlap with known gene locations.Claverie J (2005). "Fewer genes, more noncoding RNA." *Science* **309** (5740): 1529–30. doi:10.1126/science.1116800. PMID 16141064.
- [9] "...the proportion of small (50-100 bp) segments in the mammalian genome that is under (purifying) selection can be estimated to be about 5%. This proportion is much higher than can be explained by protein-coding sequences alone, implying that the genome contains many additional features (such as untranslated regions, regulatory elements, non-protein-coding genes, and chromosomal structural elements) under selection for biological function." Mouse Genome Sequencing Consortium (2002). "Initial sequencing and comparative analysis of the mouse genome." *Nature* **420** (6915): 520–62. doi:10.1038/nature01262. PMID 12466850.
- [10] The ENCODE Project Consortium (2007). ""Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project"". *Nature* **447**: 799–816. doi:10.1038/nature05874.
- [11] Zhandong Liu, Santosh S Venkatesh and Carlo C Maley, *Sequence space coverage, entropy of genomes and the potential to detect non-human DNA in human samples*, BMC Genomics 2008, 9:509, doi:10.1186/1471-2164-9-509 (<http://www.biomedcentral.com/1471-2164/9/509>), fig. 6, using the Lempel-Ziv estimators of entropy rate.
- [12] at Project Gutenberg (<http://www.gutenberg.org/browse/authors/h#a856>), zipped plaintext file, includes header information
- [13] "PLoS Biology: A New Human Genome Sequence Paves the Way for Individualized Genomics" (<http://biology.plosjournals.org/perlserv/?request=get-document&doi=10.1371/journal.pbio.0050266&ct=1>). Biology.plosjournals.org. doi:10.1371/journal.pbio.0050266. . Retrieved 2009-05-31.
- [14] Updated Dec.5,2008 08:25 KST (2008-12-05). "Digital Chosunilbo (English Edition) : Daily News in English About Korea" (<http://english.chosun.com/w21data/html/news/200812/200812050003.html>). English.chosun.com. . Retrieved 2009-05-31.
- [15] "Kbs World" (http://rki.kbs.co.kr/english/news/news_science_detail.htm?no=3623). Rki.kbs.co.kr. 2008-12-11. . Retrieved 2009-05-31.
- [16] Harmon, Amy (2008-03-04). "Gene Map Becomes a Luxury Item" (<http://www.nytimes.com/2008/03/04/health/research/04geno.html>). *The New York Times*. . Retrieved 2010-04-02.

- [17] "Human Genome Project Completion: Frequently Asked Questions" (<http://www.genome.gov/11006943>). genome.gov. . Retrieved 2009-05-31.
- [18] "What's a Genome?" (http://www.genomenewsnetwork.org/resources/whats_a_genome/Chp3_1.shtml). Genomenewsnetwork.org. 2003-01-15. . Retrieved 2009-05-31.
- [19] NCBI_user_services (2004-03-29). "Mapping Factsheet" (<http://www.ncbi.nlm.nih.gov/About/primer/mapping.html>). Ncbi.nlm.nih.gov. . Retrieved 2009-05-31.
- [20] "About the Project" (<http://www.hapmap.org/abouthapmap.html>). HapMap. . Retrieved 2009-05-31.
- [21] "2008 Release: Researchers Produce First Sequence Map of Large-Scale Structural Variation in the Human Genome" (<http://www.genome.gov/27026113>). genome.gov. . Retrieved 2009-05-31.
- [22] Nature. "Mapping and sequencing of structural variation from eight human genomes : Abstract" (<http://www.nature.com/nature/journal/v453/n7191/abs/nature06862.html>). Nature. doi:10.1038/nature06862. . Retrieved 2009-05-31.
- [23] from Bill Clinton's 2000 State of the Union address (<http://clinton4.nara.gov/WH/SOTU00/sotu-text.html>)
- [24] Nature. "Global variation in copy number in the human genome : Article : Nature" (<http://www.nature.com/nature/journal/v444/n7118/full/nature05329.html>). Nature<!. . Retrieved 2009-08-09.
- [25] Online Mendelian Inheritance in Man (OMIM) (<http://www.ncbi.nlm.nih.gov/Omim/mimstats.html>)
- [26] "Our results imply that humans and chimpanzees differ by at least 6% (1,418 of 22,000 genes) in their complement of genes, which stands in stark contrast to the oft-cited 1.5% difference between orthologous nucleotide sequences." Demuth J, De Bie T, Stajich JE, Cristianini N, Hahn MW (2006). "The Evolution of Mammalian Gene Families.". *Plos One* **1** (1): E85. doi:10.1371/journal.pone.0000085. PMID 17183716.
- [27] "We calculate the genome-wide nucleotide divergence between human and chimpanzee to be 1.23%, confirming recent results from more limited studies." The Chimpanzee Sequencing and Analysis Consortium (2005). "Initial sequence of the chimpanzee genome and comparison with the human genome.". *Nature* **437** (7055): 69–87. doi:10.1038/nature04072. PMID 16136131.
- [28] "Human chromosome 2 resulted from a fusion of two ancestral chromosomes that remained separate in the chimpanzee lineage" The Chimpanzee Sequencing and Analysis Consortium (2005). "Initial sequence of the chimpanzee genome and comparison with the human genome.". *Nature* **437** (7055): 69–87. doi:10.1038/nature04072. PMID 16136131.
"Large-scale sequencing of the chimpanzee genome is now imminent."Olson M, Varki A (2003). "Sequencing the chimpanzee genome: insights into human evolution and disease.". *Nat Rev Genet* **4** (1): 20–8. doi:10.1038/nrg981. PMID 12509750.
- [29] "Our findings suggest that the deterioration of the olfactory repertoire occurred concomitant with the acquisition of full trichromatic color vision in primates." Gilad Y, Wiebe V, Przeworski M, Lancet D, Pääbo S (2004). "Loss of olfactory receptor genes coincides with the acquisition of full trichromatic vision in primates.". *PLoS Biol* **2** (1): E5. doi:10.1371/journal.pbio.0020005. PMID 14737185.
- [30] Sykes, Bryan (2003-10-09). "Mitochondrial DNA and human history" (http://genome.wellcome.ac.uk/doc_WTD020876.html). The Human Genome. . Retrieved 2006-09-19.
- [31] "Misteli" (<http://www.cell.com/content/article/abstract?uid=PIIS0092867407001262>). Cell. doi:10.1016/j.cell.2007.01.028. . Retrieved 2009-05-31.
- [32] "Bernstein et al." (<http://www.cell.com/content/article/abstract?uid=PIIS0092867407001286>). Cell. doi:10.1016/j.cell.2007.01.033. . Retrieved 2009-05-31.
- Carroll, Sean B. et al. (May 2008). "Regulating Evolution". *Scientific American*. New York: Scientific American Inc.. pp. 60–67.
- Lindblad-Toh K, et al. (2005). "Genome sequence, comparative analysis and haplotype structure of the domestic dog." (<http://www.nature.com/nature/journal/v438/n7069/abs/nature04338.html>). *Nature* **438** (7069): 803–19. doi:10.1038/nature04338. PMID 16341006.

External links

- The National Human Genome Research Institute (<http://www.genome.gov/>)
- Ensembl (<http://www.ensembl.org/>) The Ensembl Genome Browser Project
- National Library of Medicine human genome viewer (http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?taxid=9606)
- UCSC Genome Browser (<http://genome.ucsc.edu/>).
- Human Genome Project (http://www.ornl.gov/sci/techresources/Human_Genome/project/info.shtml).
- Sabancı University School of Languages Podcasts What makes us different from chimpanzees? by Andrew Berry (<http://www.sabanciuniv.edu/do/eng/PodCast/files/podcast18.mp3>) (MP3 file)
- The National Office of Public Health Genomics (<http://www.cdc.gov/genomics/default.htm>)
- New findings challenge established views about human genome (<http://www.spiritindia.com/health-care-news-articles-10638.html>)

Article Sources and Contributors

DNA *Source:* <http://en.wikipedia.org/w/index.php?oldid=355095288> *Contributors:* (:, (jarbar), -Majestic-, 168..., 168..., 169, 17Drew, 3dsceince, 4u1e, 62.253.64.xxx, 7434be, 84user, A D 13, A bit iffy, A-giau, Aaaxlp, Aatomic1, Academic Challenger, Acer, Adam Bishop, Adambiswanger1, Adamstevensun, Adashiel, Adenosine, Adrian.benko, Ahoerstemeier, Aitias, Aj123456, Alai, Alan Au, Aldaron, Aldie, Alegoo92, Alexandremas, Alkivar, Alphachimp, Alzhaid, Amboo85, Anarchy on DNA, Ancheta Wis, AndonicO, Andre Engels, Andrew wilson, Andreww, Andrij Kursetsky, Andycjp, Anita1988, Anomalocaris, Antandrus, Ante Aikio, Anthere, Anthony, Anthony Appleyard, Antilived, Antony-22, Aquaplus, Aquilla34, ArazZeynili, Arcadian, Ardyn, ArielGold, Armored Ear, Artichoker, Asbestos, Astrowob, Atlant, Aude, Autonova, Avala, AxelBoldt, AySz88, AzaToth, B, BD2412, BMF81, Banus, BaronLarf, Bbatsell, Bci2, Bcorr, Ben Webber, Ben-Zin, BenBildstein, Bender235, Benjah-bmm27, Bensaccount, Bernie Sanders' DNA, Bevo, Bhadani, Bhar100101, BiH, Bijee, BikA06, Bill Nelson's DNA, Billmcgn189, Biolinker, Biriwilg, Bjwebb, Bkell, Blastwizard, Blogger, Blondtraillite, Bmtbomb, Bobblewik, Bobo192, Bongwarrior, Borisblue, Bornhj, Brian0918, Brighterorange, Briland, Brim, Brockett, Bryan, Bryan Derksen, CWY2190, Cacycle, Caerwine, Cainer91, Cal 1234, Calaschysm, Can't sleep, clown will eat me, Canadaduane, Carbon-16, Carcharoth, Carlo.milanesi, Carlwev, Casliber, Cathalgarvey, CatherineMunro, CattleGirl, Causa sui, Cburnett, Cerberus lord, Chanora, Chanting Fox, Chaojoker, Charm, Chill Pill Bill, Chino, Chodges, Chris 73, Chris84, Chuck Grassley's DNA, Chuck02, Clivedelmonte, ClockworkSoul, CloudNine, Collins.mc, Colorajo, CommonsDelinker, Conversion script, Cool3, Coolawesome, Coredesat, Cornacchia123, Cosmotron, Cradlelover123, Crazycomputers, Crowslab, Crusadeonlinearity, CryptoDerk, Crzrussian, Cubsrazy29, CupOBeans, Curedh, Curps, Cyan, Cyclonenim, Cyrius, D6, DIREKTOR, DJAX, DJRafe, DNA EDIT WAR, DNA is shyT, DVD R W, Daniel Olsen, Daniel987600, Danielkueh, Danny, Danny B-), Dansk14, Darklark, Darth Panda, Davegrupp, David D., David Eppstein, Davidbspalding, Daycd, Db099221, Dbabbitt, Doceetee, DeAceShooter, DeadEyeArrow, Dellidot, Delta G, Deltabaignet, DevastatorIIC, Diberri, Dicklyon, Digger3000, Digitalme, DNA, Djm1279, Dlohcierekim's sock, Dmnn, DocWatson42, Docjames, Doctor Faust, Docu, Dogposter, DonSiano, Donareiskoffer, Dr d12, Dr.Kerr, Drini, Dudewhereismywallet, Dullhunk, Duncan.france, Dungleodung, Dysmorodrepanis, E. Wayne, ERcheck, ESKog, Echo park00, Echuck215, Eddycrai, Editing DNA, Edwy, Efbfweborg, Egil, ElTyran, Elb2000, Eleassar777, EliasAlucard, ElinorD, Ellmist, Eloquence, Emotionc, Emw, Ephemerionum, Epingehris, Erik Zachte, Escape Artist Swyer, Esumir, Etanol, Ettrig, EurekaLott, Everyking, Evil Monkey, Ewawer, Excavator, FOTEMEH, Fabhcn, Factual, Fagstein, Fastfission, Fconaway, Frick, Fernando S. Aldado, Firehorse, Figma, Figure, Firetrap9254, Fishingpal99, Flavaflav1005, Florentino floro, Frielsen, Forlufvof, Freakofnurture, FreplySpang, Friendly Neighbour, Frostyservant, Fruge, Fs, Fvascencellos, G3pro, GAThrawn22, GHe, GODhack, Gaara san, Galoubet, Gary King, Gatortpk, Gazibara, GeeJo, Gene Nygaard, GeoMor, Giftlite, Gilisa, Gilliam, Gimmetrow, Gjuggler, Glen Hunt's DNA, Glenn, Gmaxwell, GoEThe, Goatsaur, Gogo Dodo, Gohnafototahabadi, GordonWatts, Gracenotes, Graeme Bartlett, Graemel., Grafikm Jr, Graft, Graham87, GrahamColm, Grandegrandegrande, GregorB, Grover cleveland, Gurko, Gustav von Humpelschmumpel, Gutza, Gwsrinme, Hadal, Hagerman, Hairchrm, Hairwheel, Hammersoft, Hannes Röst, Harianto, HayleyJohnson21, Heathunnicutt, Hephaestos, Heron, Heyheyhack, Hockey21dude, Horatio, Hu, Hughdbrown, Hurricanehink, Hut 8.5, Hvn0413, I hate DNA, Iapetus, Icairms, Ilia Kr., Impamizigraa, InShanee, Inge-Lyubov, Isilan, Isis07, Itub, Ixfd64, Izehar, JHMM13, JWSchmidt, JWSurf, Jacek Kendysz, Jaccorn, JamesMLane, JamesMtl1984, Janejellyroll, Javert, Jaxl, Jax9111, Jer ome, JeremyA, Jerzy, Jetssetpainter, Jh51681, Jiddisch, Jimriz, Jimwong, Jlh29, Jls043, Jmcc150, Jo9100, JoanneB, Joconnol, Joeywallace9, Johanvs, JohnArmagh, Johtex, Johnuniq, Jojit fb, JonMoulton, Jonrunles, Jonverve, Jorvik, JoshuaZ, Josq, Jossi, Jstech, Julian Diamond, Jumbo Snails, Junes, Jwrosenzweig, Kahlfm, Kapow, Karrmann, Kazkaskazkasako, Khb3rd, Keegan, Keepweek, Keilana, Kelly Martin, Kemyou, Kendrick7, KerryO77, Kevin B12, Kevmitch, Kgheos, Kholdstare99, Kieran, KimvdLinde, King of Hearts, KingTT, Kingturtle, Kitch, Knowledge Seeker, KnowledgeOfSelf, Koavf, KrakatoaKatie, Kums, Kungfuadam, Kuru, Kwamikagami, Kwekubo, KyNephi, LA2, LFaraoe, La goutte de pluie, Lascorz, Latka, Lavateraguy, Lee Daniel Crocker, Lemchesvej, Leptictidium, Lerdsuwa, Leuko, Lexor, Lhenslee, Lia Todua, LightFlare, Lightmouse, Lightspeedchick, Liguem, Lincher, Lion Wilson, Lir, Llongland, Llull, Lockesdonkey, Logical2u, Loginbudwe, Looxix, Loren36, Loris, Luigi30, Luk, Lumos3, Luna Santin, Luuva, MER-C, MKoltnow, MONGO, Mac, Madeleine Price Ball, Madhero88, Magadan, Magnus Manske, Majorly, Malcolm rowe, Malo, Mandyj61596, Mantissa128, Marcus.aerlous, Marj Tiefert, Martin.Budden, MarvPaule, Master dingley, Mattbr, Mattbrundage, Mattjbylte, Maurice Carbonaro, Mav, Max Baucus' DNA, Max Naylor, McDogn, Medessec, Medos2, Melaen, Melchoir, Mentalmania07, Mgitanteus1, Mgttoehy, Mhking, Michael Devore, MichaelHa, Michaelas10, Michigan user, MidgleyDJ, Midnightblueowl, Midoriko, Mika293, Mike Rosoff, Mikker, Mikko Paananen, Mintman16, MisfitToys, Miszal31, Mithent, Mjpieters, Mleefs7, Moink, Moorice, Mortene, Mr Bungle, Mr Meow Meow, Mr Stephen, MrErku, Mstislavl, Mstrocek, Mulad, Munita Prasad, Muro de Aguas, Mwanner, Mxn, Nakon, Narayanese, Natalie Erin, Natarajanganesan, Nate1028, NatureA16, Nauseam, Nbauman, Neckro, Netkinetic, Netoholic, Neutrality, Never give in, NewEnglandYankee, Nighthawk380, NighthawkJ, Nihiltres, Nirajrm, Nishkid64, Nitecrawler, Niteamreknap, No Guru, NoldeaNick, NochnoiDozor, Nohat, Northfox, NorwegianBlue, Nthornberry, Nunh-huh, ObloodHell, OOODDD, Obli, Oblivious, Ocolon, Ojl, Omicronperse8, Onco p53, Opabinia regalis, Opelio, Orrin Hatch's DNA, Orthologist, Ortolan08, Ouishoebean, Outriggr, OwenX, P99am, PDH, PFHLai, PaePae, Pakaran, Pascal666, Patrick0Moran, Patrick2480, Patstuart, Paul Foxworthy, Paul venter, Paulinho28, Pcb21, Pde, Peak, Pedro, Persian Poet Gal, Peter Isotalo, Peter K., Peter Winnberg, Pgan002, Philip Trueman, PhilipO, Phoenix Hacker, Pierceno, PierreAbbat, Pigman, Pigmiethelclub, Piloguy, Pkirin, Poor Yorick, Potnue6927, Potatoswatter, Preston47, Priscilla 95925, Pristoncalt111, Pro crast in a tor, Prodego, Psora, PsvMar, Pysmier, Pumpkingrower05, Pyrospirrit, Quebec99, Quickbeam, Qutezuce, Qxz, R'n'B, R. S. Shaw, RDBrown, RJC, RSido, Ragesoss, Rajwiki123, RandomP, Randomblue, Raul654, Raven in Orbit, Ravidreams, Rdb, Rdsmith4, Red Director, Reddi, Rednblu, Redneckjimmy, Redquark, Retired username, Retteast, RexNL, Rich Farmbrough, RichG, Richard Durbin's DNA, Ricky81682, Rjwilmsi, Roadnottaken, Robdurbar, RobertG, Rocastelo, RoddyYoung, Rory096, Rotem Dan, Roy Brumback, RoyBoy, RoyLaurie, Royboughard111, RunOrDie, Russ47025, Rxs, RyanGerbil10, Ryulong, S77914767, SCEhard, STAN SWANSON, SWAdair, Sabbage, Safwan40, Sakkura, SallyForth123, Sam Burne James, Samsara, Samuel, Samuel Blanning, SandyGeorgia, Sango123, Sangwine, Savidan, Scarce, Sceptre, Schutz, Sciencechick, Scienceman123, Scinciesociety, Sciuirina, Scope creep, Scoterician, Sean William, SeanMack, Seans Potato Business, SebastianHawes, Seldon1, SemperBlotto, Sentausa, Serphine, Shadowlynk, Shanes, ShaunL, Shekharsuman, Shizhao, Shmee47, Shoy, Silsor, SimonD, Sintaku, Sir.Loin, Sjjupadhyay, Sjollema, Sloth monkey, Slrubenstein, Sly G, SmilesALot, Smithbrenon, Snowmanradio, Snowolf, Snurks, Solipsist, Someone else, Sonett72, Sopoforic, Spaully, Spectrogram, Spiff, Splette, Spondoolicks, Spongebobsqpants, SpuriousQ, Squidonius, Squirepants101, Statsone, Steel, Steinsky, Stemonitis, Stephen, SteveHopson, Stevertigo, Stevietheman, Stewartadcock, Stuart7m, Stuhacking, SupaStarGirl, Supspirit, Susvolans, Sverdrup, Swid, Switchercat, Taco325i, Takometer, TakuyaMurata, Tariqabjotu, Tarret, Taulant23, Tavilis, Tazmaniacs, Ted Longstaffe, Tellyddick, TenOfAllTrades, Terraguy, Test100000, TestPilot, The Rambling Man, The WikiWhippet, TheAlphaWolf, TheChrisD, TheGrza, TheKMan, TheRanger, Thorwald, ThreeDaysGraceFan101, Thue, Tiddly Tom, Tide rolls, TigerShark, TimVickers, Timewatcher, Timir2, Timl2k4, Timrollpicking, Timwi, Tobogganogin, Toby Bartels, TobyWilson1992, Tom Allen, Tom Harkin's DNA, Tomgally, Toninu, Tony1, Tonyrenpliki, Trd300gt, Trent Lott's DNA, Triwbe, Troels Arvin, Tstrobaugh, Tufflaw, Turnstep, Twilight Realm, Ty146Ty146, UBeR, Unint, Unukonome, Usergreatpower, Utcursch, Uthbrian, Vaernmond, Vandelizer, Vanished user, Vary, Virtualphtn, Visium, Vividonset, VladimirKorablin, Vsmith, Vyasa, WAS 4.250, WAVEgetarian, WHeimbigner, WJBscribe, Wafulz, WarthogDemon, Wavelength, WelshMatt, West Brom 4ever, Where, Whosasking, Whoutz, Why My Fleece?, Wik, Wiki alf, Wiki emma johnson, Wikiborg, Wikipedia Administration, William Pietri, WillowW, Wimt, Wknight94, Wmaham, Wnt, Wobble, WolfmanSF, Wouterstomp, Wwwwwolf, Xy7, YOUR DNA, Yamamoto Ichiro, Yama, YanWong, Yansa, Yaser al-Nabriss, Yasha, Yomama9753, YOUNUSPORTEOUS, Yurik, ZScout370, Zahid Abdassabur, Zahiri, Zanaq, Zazou, Zell Miller's DNA, Zephyris, Zoicon5, Zouavman Le Zouave, Zsinj, Zven, 1329 anonymous edits

DNA sequence *Source:* <http://en.wikipedia.org/w/index.php?oldid=354763602> *Contributors:* 168..., Abecedare, Al Lemos, Aranae, Bealegotpinky, Diagonalfish, Drgarden, Drphilharmonic, El Raki, ForestDim, Forlufvof, Frank Lofaro Jr., Gallan, Gimboid13, Hugo-cs, I am fl, Jamesters, Jaxl, Johnuniq, Jojhutton, Kku, La goutte de pluie, Lexor, LiHelpa, LittleOldMe, Michael Hardy, Mushiophil, Nagelfar, Narayanese, Netopaa, Peak, Peregrine Falcon, Physicistjedi, Ragesoss, Reach Out to the Truth, Robofish, Rosarinagazo, RoyBoy, Shoefdeath, Shureg, Stemonitis, SteveChervitzTrutane, TheNoise, Ultimate sickness, Wavelength, Wli625, 40 anonymous edits

DNA sequencing *Source:* <http://en.wikipedia.org/w/index.php?oldid=352971905> *Contributors:* 25orf6t04, 5piggies, ANSMID, Abanima, Abdualah mohammed, Abdull, Abizar, Ace of Spades, Adrian J. Hunter, Agricola44, Akita86, Akjosh, Akriasas, Anradt, AxelBoldt, BaChev, Bhar100101, Bjorn9800991, Brice one, Charles Gaudette, Cinnamon colbert, Clicketyclack, ClockworkSoul, ConfuciusOrnis, Cpeditorial, Crana, Cremepuff222, Cspenser, CultureDrone, DWeissman, Dai mingjie, David Schaich, DerHexer, Discospinster, Dkropf, Dnacash, DoctorDNA, Drcsk, Duncan, Dupz, DutchDevil, Dysmorodrepanis, ERcheck, Eimeim, Ejwong, Enzymes-GMU, Eric-Westar, Excirial, Eyu100, Fedra, Frankatca, Furby100, Fuxx, Genomemer, George Church, God Emperor, Graminophile, HappyInGeneral, Helwr, Hoe4sho, Howie Goodell, Jacopo Werther, Johnuniq, Jvbishop, Kadoo, Keaka411, Ketil, Kku, Kymacpherson, Lantonov, Lenoxus, Lmsutton1, Loris, Luckas Blade, MBCF, Madeleine Price Ball, Magnus Manske, Malljajia, MarcoTolo, Marscher23, Mikael Häggström, Mikestork, Minima's Clone, MrOllie, Naturespace, NawlunWiki, Neffk, NewEnglandYankee, Nono64, Novangelis, Noveltyghost, Ohnoitsjamie, Omeatron, Oxyron83, PDH, Patho, Peripitus, Petter Bockman, Pharos, Pixelface, PrincessofLjyr, Pvosta, RDBrown, RJaguar3, RRphys, Red89011, Rees11, Res2216firestar, Rich Farmbrough, Rjwilmsi, Robert K S, Robin klein, Robinatron, Rosyaraur, RoyBoy, SJP, SNP, Sandahl, Schnarr, Schtina, Seans Potato Business, Sentausa, Sheha, Shrimp wong, Shyamal, Sintaku, Sirozha, SomethingCatchy, Spellmaster, SteveChervitzTrutane, Su-no-G, Svick, Tammyz06, TestPilot, Thisisborin9, Tide rolls, Trevyn, Tsmith423, Twas Now, Victor D, Vinylmeister, Wavelength, Webpolish, Xenon54, Xook1kai Choa6aur, Yurivict, Zaimon, 217 anonymous edits

DNA profiling *Source:* <http://en.wikipedia.org/w/index.php?oldid=354964979> *Contributors:* 123imran, 168..., 7434be, 808OG33, ABF, Abberley2, Abeg92, Abrahami, Abune, Acegikmo1, Aeternum8, Ahoerstemeier, Aitias, Ajraddatz, Akendall, Alansohn, Alexbrewer, Allstarecho, Alphachimp, Angelo De La Paz, Anneman, Anonymous editor, Anyo Niminus, Ap, Apers0n, Appraiser, ArchonMagnus, Arcturus, Asyndeton, Aude, Awostrack, AxelBoldt, AzaToth, BarrelBonden, Bastiche, Beetstra, Bidgee, Blaaka, Bobo192, BorisTM, Bradjamesbrown, Brandoobbe, Brumski, Bugg42, CHallam, CFreland, CNicol, Cacycle, Caltas, Can't sleep, clown will eat me, Capricorn42, Captain Obvious and his crime-fighting dog, CardinalDan, Catgut, Ceoil, CharlesBrenner, Charm, Charm, Chasingol, Chris the speller, Chrisl02, Christopherlin, Click23, Cmcenicol, C12288, Curps, DJ Clayworth, Dan100, Danorton, DavidWBrooks, DeadEyeArrow, DeBresser, Dharmabum420, Discospinster, Dkropf, Dnalasindia, DocWatson42, Domikins, Doom7987, DougsTech, DragonflySixtyseven, Drunken Pirate, Dubitante, Duncharris, Dungodung, DeRahier, E Wing, ElectricLemon, Emperorbma, Enviroboy, Epr123, Eperotau, Everyking, Ewawer, F-451, FJBP, Fastfission, Feedmecereal, Fieldday-sunday, Figma, Formdadoxfraud, Freechild, Funkboy3, GIR, GRBerry, Gaius Cornelius, Galoubet, Gill110951, Gilliam, Ginsengbom, Golbez, GraemeL., GreaterWikiholic, Ground Zero, Gurch, Gwernol, HalJor, HalfShadow, HamburgerRadio, Happenstance, Hardware Hank, HarringtonSmith, Hephaestos, Hyper84, Iapetus, Icarus3, Imzadi1979, InspectorSands, Inter, Iridescent, IronGargoyle, Islaammaged126, Ixymapoe, J.delanoy, J04n, JBarno, JForget, JMS Old Al, JNW, Jake Wartenberg, James Callahan, Jamesontai, Jeffire, Jengod, Jenwen28, Jetjaxon5, Jimokay, Johnuniq, Joseph Solis in Australia, Josh Parris, Jovianeye, Jpbowen, Justin, Justpassin, Karada, Keepmythings, Kembangraps, Kingpin13, KnightRider, Koffieyahoo, Ksryie, Kupirijo, Kuru, Kvdever, LachlanA, Lcmtuf, LegendGamer, Lexor, Lightmouse, Linnhall, LizardJr8, Lidancer91, Lonesomefighter, MER-C, MONGO, MacsBug, Macy, Madeleine Price Ball, Mandarax, Manuel Yayo, Marandbon, MarcoTolo, Markbencke, Marketing250, Martin.Budden, Marysunshine, Materialsscientist, Mattisie, Mav, Maverick Leonhart, Maximus Rex, McPavement, MelchiorZ, Mercury, Mervyn, Mgalie, Muggmess, Michael Devore, Michael Hardy, Michal Nebyla, Mikael Häggström, Mindmatri, Momoricks, Motor.on, NHRHS2010, Naive cynic, Nightscream, Nitroshockwave, Noahsachs, Nuggetboy, Obli, OllieFury, Oore, Pakaran, PaleWhaleGail, Paternity Test, Patrick, Patstuart, Paulbrock, Pengo, Personman, Peruvianllama, Peter johnson4, Petiail,

Pgk, Phantomsteve, Phil Holmes, Philbarker, Piccadilly, PleaseStand, Poindexter Propellerhead, Prodego, Propaniac, PseudoOne, Puchiko, Pv86, Quadell, Qxz, RadioFan2 (usurped), RattusMaximus, Ravichandar84, Red Director, Reinyday, Rettetast, Rich Farmbrough, Richwales, Rjwilmsi, Robth, Ron Barker, Rorro, RoyBoy, Rsaanie, Rugbyboroughman, Russ47025, S-MorrisVP, Sandahl, Sandstein, Sarah00001, SardanaPhalus, Schutz, Scubafish, Sekiyu, Sens08, Serpens, Severa, Shaddock, ShakingSpirit, Shimeru, Shumdw, Skidude9950, Skipper500, Skyk98, SlowJog, Smartse, Smor4860, Snicklefritzdee, SnowsCode, SoLando, Soarhead77, Someguy1221, Someone else, SpaceFlight89, Spencer BOOTH, Spongeboblover124, Steinsky, Stepa, Steven Zhang, StuartMcIntyre, Stubbyhead, SyntaxError55, Syrrhis, TBadger, TaintedMustard, Tameeria, Tamoxidiva, Tanthalas39, Tazmaniacs, TedE, TenOfAllTrades, TexasAndroid, ThatguyTint, The Baroness of Morden, The wub, TheFreeelador, Thincat, Think outside the box, Tiddly Tom, Tide rolls, TimVickers, Tioxod, Tizio, Toytown Mafia, Tresiden, Tstrobaugh, Twang, Uksam88, Uncke Herb, Uncle Dick, Universalgenetics, VernoWhitney, Versus22, Vicenarian, Vifsm, Viriditas, Vsmith, WLRoss, Wavelength, Werdan7, Wiki alf, Wikid77, Wjcollier07, Woohookitty, Wwc93, Xavierschmit, Yahel Guhan, YanWong, Yerro, Zashaw, Zophra, Zé da Silva, Bj666lfr, आशीष भट्टनागर, 835 anonymous edits

DNA polymerase *Source:* <http://en.wikipedia.org/w/index.php?oldid=355235163> *Contributors:* 168...., 84user, Abendwind, Abstraktn, Accurizer, Afasmit, Agathman, Amazinglarry, Arcadian, Aziz1005, Bensaccount, Billare, Bomac, BorisTM, Brutha, Cafall42, Christopherlin, Classof2000, ClockworkSoul, DerHexer, Drphilharmonic, Dullhunk, Eisor, Emersoni, Emw, EnDumEn, Eprb123, Erik Garrison, FriedMilk, G3pro, GiollaUdir, Gorton k, Grafen, Graham87, Hannes Röst, Hede2000, Heron, Improv, Jag123, Jimhsu77479, Jjwarren, John Cumbers, John.nelson@ge.com, Julianonions, Jusjih, KBi, Kaarel, Kanags, Kazkaskazkasako, Keesiewonder, Kuru, Kwc148, Leevanjackson, Leuko, Lxor, LovesMacs, MER-C, Madeleine Price Ball, Magnus Manske, Maximus Rex, Mcbzac, Michael Snow, Mikael Häggström, Mineralogy, Modulatum, Musescooty, Napster123456789, Nick, Nneonneo, Noahsachs, Ojigiri, Oneiros, PDH, Pner Znamenskiy, Pinnerup, Polluks, Rdsmith4, RelentlessRecusant, Rheka, Rigadoun, Rror, Saffile, Shrimp wong, Slicky, Srootsel, Taka, Tanthalas39, The Thing That Should Not Be, TheTwiz, TimVickers, Tycho, Velela, Vojtech.dostal, Whosasking, Will Wh, Yahel Guhan, Ychastnik APL, Youssefsan, 이형주, 184 anonymous edits

DNA Topoisomerase *Source:* <http://en.wikipedia.org/w/index.php?oldid=345472715> *Contributors:* 168...., Alexhlau, Anthony Appleyard, Antorjal, Arcadian, Axl, Banus, Berkay0652, Boghog2, Capecodeph, CarTick, Chochopk, Chris the speller, Chrumps, Clicketyclack, ClockworkSoul, DabMachine, DanaStreet, Dauto, Discospinster, East718, EquationDoc, Forluvoft, G3pro, Gharmon, Giftlite, Greg carter, Gurch, Intrgr, Jafield, Japanese Searobin, Jfdwelf, Kcordina, Kunalmehta, Linefeed, Lorangriel, Luuva, MBisanz, Miguel Andrade, Nbauman, Orthank, Passw0rd, Pdeitiker, Pixelface, Robma, SGGH, SenorKristobbal, Shyamal, Stemonitis, Stevenjohnkiss, TenOfAllTrades, Thorwald, Ummel, Vojtech.dostal, Whiner01, Wikididact, Youssefsan, 82 anonymous edits

List of nucleic acid simulation software *Source:* <http://en.wikipedia.org/w/index.php?oldid=350978652> *Contributors:* Ataricom, Colonies Chris, Cureden, Dawynn, Enric Naval, Michael Hardy, P99am, Thorwald, Vizbi, WikHead, 2 anonymous edits

DNA structure *Source:* <http://en.wikipedia.org/w/index.php?oldid=355209793> *Contributors:* Antony-22, Shadowjams

Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid *Source:* <http://en.wikipedia.org/w/index.php?oldid=352349710> *Contributors:* Abrech, Addshore, Afreinside13t, Bci2, Bkbbrrad, Bubbachuck, Camw, Carcharoth, Casey Abell, ChazYork, ClockworkSoul, CommonsDelinker, Danshil, DerHexer, Easayer, Eldaran, Erkan, Fratper, GiscaCobra, Grenavitar, JWSchmidt, Jacek FH, Jacob Finn, Jnc, John Vandenberg, Kaarel, Koavf, La goutte de pluie, LilHelpa, Lizzie Harrison, MFH, Matthew Yeager, Maxis ftw, Mithent, Mushin, N6, Neilio 23, Northfox, Nurasko, Oxymoron83, Patrick, Petersec, Petri Krohn, RJaguar3, Rich Farmbrough, RobertG, SMA11784, Sam Korn, Schnarr, Selket, Shenme, Snalwbma, Snapperman2, Someguy1221, Swid, Switchercat, Tempodivalse, Thorwald, Thunder, Tiddly Tom, Tim!, Willking1979, WongFeiHung, Wouterstomp, Zzuuzz, 113 anonymous edits

Molecular models of DNA *Source:* <http://en.wikipedia.org/w/index.php?oldid=355339635> *Contributors:* Antony-22, Bci2, Chris the speller, CommonsDelinker, David Epstein, Jwoodger, Michael Hardy, Miym, Nmedard, Oscartheat, P99am, Tassedethe, The Thing That Should Not Be, 2 anonymous edits

DNA replication *Source:* <http://en.wikipedia.org/w/index.php?oldid=354378607> *Contributors:* 101louise101, AKaK, Adambro, Adenosine, Aesopos, Agathman, Ahoerstemeier, Aitias, Alansohn, Alfio, Alphachimp, Alveolate, Andthendousaid, Andy M. Wang, Arcadian, Archdevil75, Ariliand, Arthmelow, Artman40, Ary29, Ashermadan, Azcolvin429, Baldhur, Bensaccount, Bio-queen, Biomedeng, Blanchardb, Blondtraillite, Bobo192, BokicaK, Bomac, Borgx, Bowlhover, Bozartas, BradBeattie, Brianwatson94, Burner0718, Cleland, CTZMSC3, Calito000000001, CanadianLinuxUser, Canis Rufus, Captain Infinity, Catgut, Celamor, Chasturn, Chris Capoccia, Cless Alvein, ClockworkSoul, Cmdrjameson, CommonsDelinker, Conversation script, Coveman999, Crazysunshineboy, Cyktsui, Dac04, Dan Gluck, Dantheman531, Dave Runger, David D., Debresser, Dennis Valeev, DerHexer, DividedByNegativeZero, Dr. Anton Funk PhD, Dusanman1, Duxenaz, Dysepion, ERcheck, Ed Fitzgerald, Edgar181, Edwy, Egania, El C, Emw, Eprb123, Erick880, Exarion, Falconus, Faradayplank, Farosdaughter, Fbv65edel, Fibonacci, Firelemont85, ForestDim, Freedomlinus, FriedMilk, GD, GLaDOS, Gail, Giftlite, Gobonobo, Golfandme, Graft, Gregoots, Guelphie, Guelphie85, Habj, Hadal, Harryboyles, Harshal112, Hello32020, Heron, Hichris, Hu12, Iduggin, Ignacio Icke, Im emo tastic, Imjustmatthew, InvaderJim42, J.delanoy, JDCMAN, JForget, Jackhynes, JadtNr1, JayZ, Jcwf, Jessi bob, Jhinman, Jimhsu77479, JoanneB, Johnuniq, Jonemerson, Jossi, Jovianeye, Jpeguero, JulianBlow, Julianonions, Jullag, Kaisershatner, Kakofonous, KnowledgeOfSelf, Kusb, LLLDMart, La goutte de pluie, LadyOfHats, Leafyplant, Lxor, LoneSeeker, Lotez, Lupo, Macduy, Madeleine Price Ball, Majorly, Marek69, Maristoddard, Maurice Carbonaro, Medfreak, Media1312, Megaman en m, Meno25, Mentifisto, Mgianteus1, MichaelHa, Midgrid, Mietchen, Mikael Häggström, Monkey Bounce, Mortice, MrErku, Mstuddert, Mufka, Mygerardromance, Narayanese, Netkinetic, Neutrality, Neverquick, Nhandler, Noah Salzman, Nonagonal Spider, Nornour, Nutriveg, Oddharmonic, Olavrg, Omes, Oxymoron83, Pb30, Persian Poet Gal, Pfranson, Pinethicket, Pkakis, Player 03, Plrk, Ppgardne, Pranay Biswas, PseudoSudo, Pstanton, RUL3R, Radon210, Raven4x4x, Ravenhull, RelentlessRecusant, Residenthalo, Rettetast, Rich Farmbrough, Richard001, Robinhaw, Romanticcynic, RoyBoy, Royalguard11, RyanGerbil10, Rystic, Sakkura, Samuel, Sandymit, Scientizzle, Sedmic, Shajilvt, Skysmith, Smalljim, Snafflekid, Snowolf, SparrowsWing, SpuriousQ, Squidonius, Steipe, Stepa, Stillnotel, Sturm55, Stwalkerster, Superandomness, Svetlana Miljkovic, Sylent, Taka, Tameeria, Tassos Kan., TastyPoutine, TexMurphy, The Anome, The Thing That Should Not Be, Thegathering, Thiseye, Thisismikesother, Tide rolls, TimVickers, Tomjc, Tree Biting Conspiracy, Tstrobaugh, Twas Now, Twooars, Tycho, Ucanlookitup, Uphilsc, Vanderesch, Victor Blacus, Vojtech.dostal, Voyagerfan5761, Vsmith, WAVEgetarian, Wangi, WaysToEscape, WebHamster, Whiteniko, Whosyourjudas, Work permit, Wouterstomp, Wrlampe, Yashgaroth, Yerro, Yuxaku, Zephyris, ZimZalaBim, Zrulli, 754 anonymous edits

DNA repair *Source:* <http://en.wikipedia.org/w/index.php?oldid=351286275> *Contributors:* ALargeElk, Abstraktn, Agathman, Alansohn, Aleenf1, Alextrevelian 006, Ali Raza, Alnokta, Amazinglarry, Amboo85, Angela, Angusmcclellan, Ariliand, Arup Acharjee, Ashbagwat, Asenine, Athf1234, Auk nambiar2002, Autrijus, Axl, Bavan Palasanthiran, Benbest, Bernstein0275, Bhuvanesh aravind, Brian0918, Brighterorange, Bruno in Columbus, CamPjxx, Cadmium, Capricorn42, Chfosli, Ching-3, Clicketyclack, Cubathy, Cyktsui, DO11.10, Dark Shikari, Dabbitt, Deepa.kushwaha, DerHexer, Deville, Discospinster, Duncharris, Ec5618, Edcolins, Edgar181, Eequor, Emvee, Emw, Eugeneltc, Fenice, FisherQueen, Frielsen, Foosh, ForestDim, Forluvoft, Gaggaj, Gerriet42, Giftlite, Ginkgo100, Gnomehacker, Gogo Dodo, GrahamColm, HaHa AI, Hadal, Harrison35, Harvestgalaxy, Herbee, Ikiroid, Ixf64, JDG, JWSchmidt, Jay Gatsby, Jimaginar, Jimktrains, JoergenB, Jon the Geek, JonHarder, Jonverve, Juliancolton, Kazkaskazkasako, Kenb215, Kiko0610, Kurtz22, La goutte de pluie, Lagnaeut, Leptictidium, Lxor, Lightmouse, LostLucidity, Maralia, Marcosantezana, Martinmmc, Mdd4696, Mike2vil, Mindmatrix, MisterSheik, MolBioMan, Monkeyman, Mr a retha, Mr. Blackout, Mushin, NYC-METS, Nina Gerlach, Ninetyone, Nunh-huh, Nutriveg, Oblivious, Opabinia regalis, Otvaltak, PAK Man, PDH, Patstuart, PervyPirate, PhiloVivero, PiKeeper, Pir, Prometheus1, Pruneau, Pschemp, Qizkajzer, RG2, Radowayne, Raul654, RebekahThorn, Red Director, Rintrah, RobertG, Robinhaw, Roncalli, Samsara, SandyGeorgia, Schneelocke, Scientist101, Serephine, Shibleoth, Snapperman2, Solipsis, Splintercellguy, Steinsky, Stephan Schulz, Stormwriter, Superdix, Synthetase, Tavisli, TimVickers, Tone, Tormsubject, Treisijs, Trilateral comm, Tualha, Tumadre36, Umarekawaru, Utcursch, Verbal, Vojtech.dostal, Wigbold, Wknight94, Wouterstomp, Yoninah, Yosri, Zafiroblue05, 224 anonymous edits

DNA Translation *Source:* <http://en.wikipedia.org/w/index.php?oldid=350988290> *Contributors:* Aciel, Agathman, Alansohn, Alexandre Vassalotti, Alnokta, Ammonfife, Antorjal, Appraiser, Arcadian, Bensaccount, BernardP, Bioinformin, Bobmarley1987, Brianwatson94, CanisRufus, Caulde, CharlesZ, Chewie, Chino, Closedmouth, Cooney118, Crum375, CrzyMke, Cunard, DB au, Dan aka jack, Dexter prog, Donimo, Duja, Dzordzm, ESKog, Eprb123, Forluvoft, Fortunecookie289, Fotinakis, Franchperez, G3pro, Gadfium, Gail, Gccwang, Giftlite, Graft, Hackwrth, Haza-w, HilJackson, Immunize, Intropy, Irrbloss, JHMM13, James McNally, Jbhood, Josh Cherry, JuanitaJP, Kaarel, Kazulanth, Kazvorpai, Kegs-cm13sx, Kim Bruning, Kostmo, Lando5, Lantonov, Lxor, Lgin, Miss1ontomars2k4, MBisanz, MBob, Marek69, Mephistophelian, MichaeJanich, Mikael Häggström, Modulatum, Mr. Know-It-All, Neokamek, Nina Gerlach, Northfox, Nwbeeson, Oldekop, Orionus, OtakuNOVAkun, PDH, PFHLai, Phe, Pro bug catcher, Psbsub, Pupunwiki, RDBrown, Rasmusw, Rolando, Saippukauppias, Salt Yeung, Sanbioinfo, Secundus Zephyrus, Shadowelvel1, SimonP, Sjkd13, SomethingCatchy, Sord halo, Squidonius, Stroppolo, Superdoggy, Sverdrup, Taoster, TenOfAllTrades, The Anome, Thisismikesother, Timemutt, Timwi, TransControl, Trinit, Uncle-P, Unfortunate, Wickey-nl, Xiahou, YaacovCR, Yashgaroth, Yasingam, Yontally, ZayFoTT, Zephyris, 230 anonymous edits

DNA Transcription *Source:* <http://en.wikipedia.org/w/index.php?oldid=353180225> *Contributors:* ASUDI, ASDZXCQWE, ATing, Agathman, Alohascott, Altzinn, Amoremeltzer, Anaxial, Andtrand, Antorjal, Apers0n, Arcadian, Archer3, ArglebargleIV, BVBed, Batmanand, Bensaccount, Berkay0652, Binksternet, Bloodpack, BokicaK, Bonás, Brettbarbaro, Brianwatson94, BrotherGeorge, Bursting74, Cbailey7, Ceyockey, CoeurDeLion, CorpX, Cvd5012, DaDrought3, DanielCD, Danjel, Dcooper, Dpv, Drosilia, Duncharris, Ec5618, EnSamuliii, Enviroboy, Epingchris, Evandrix, Fbv65edel, Feinoha, Fenoxielo, Forluvoft, Foureyes915, Frigginnacky, Furfurfor, Fuzheado, G. Völcker, G3pro, Galneon23, Gamingmaster125, Geir.overland, Geno-Supremo, Gentlemaan, Giftlite, Gillyweed, Grunt, Haham hanuka, Hede2000, Helixblue, Hilwhale, HoergerJ, Hpswimmer, Hughitt1, Hydrogen Iodide, Ifan160, Iknouyuridex, Ilia Kr., Iridescent, JForget, JWSchmidt, Jcdietz03, Jcorry10, JoanneB, John254, Johnuniq, Jullag, KDSKDS, Kaarel, Kazkaskazkasako, Khoikhoi, Kickassso, Kingpin13, Kku, Kshieh, Kubigula, Kukini, Kuru, Kzhr, Lantonov, LeaHazel, Legolost, Lehtv, LeighClesterMolar, Lxor, Liam Skoda, Local hero, LostLucidity, MDG38, MarkSutton, Marshallsumter, Martyn Axon, Mav, Maxis ftw, Mengxu, Mexican9493, Mgttoohy, Michael Devore, Michael Hardy, Miguel Andrade, Mikael Häggström, Mindmatrix, Mishlai, Mms, Moogledan, Mxn, NERIU, Narayanese, Nburden, NeurOX, Nick wiebe, NightwolfAA2k5, Nishkid64, Noah Salzman, Nubiatche, Nuplex, Onevalefan, Opabinia regalis, Opus118, OverlordKain, Oxhop, PFHLai, PGWG, Pabloes, Paulabek, Perfecto, Peter M Dodge, Phi'nLx, Phil Boswell, Postlock, Qrex123, Qwfp, R'n'B, R-skin, RaseaC, RelentlessRecusant, Rettetast, Rgocs, Riana, Ribrob, Richard001, Rmky87, Rod57, Ronz, RyanGerbil10, S73v3n, Su Nungui, Sakkura, Sameerbaab, Saurabh523, Sresawan, Seans Potato Business, Segabud, Shanell, Sjakkalle, SI, Smelissali, Sophos I, Squidonius, Sriram sh, Srlasky, Staffwaterboy, Steinsky, Teethies563, Teply, TestPilot, The Flying Spaghetti Monster, The undertow, TimVickers, Timemutt, Tmnh, Tomjc, Totophe64, Transitiveinstance, Tycho, Useight, Uthbrian, Vidric, Vikky2904, Vojtech.dostal, WAS 4.250, WAVEgetarian, WikHead, Wouterstomp, Yerro, Yided, Yvorez1274, Z.E.R.O., Zashaw, Zephyris, Περιεργος, Александър, 525 anonymous edits

DNA Transfer *Source:* <http://en.wikipedia.org/w/index.php?oldid=354696710> *Contributors:* -Midorihana-, Adrian J. Hunter, Agathman, Aircorn, AndonicO, Andrewericoleman, Arcadian, Arthana, Arup Acharjee, AxelBoldt, Bendzh, BobofBobs, Boing! said Zebedee, Bryan Derksen, Celefin, Chris the speller, Chuq, CitizenB, Cmdrjameson, Colincbn, Coolbrd, Dclyde, Dmmaus, Dr d12, Duncan.france, Ejedias, Eperotao, Excirial, Fconaway, Forluvoft, G. Völcker, Genomancer, Gharmon, Gilliam, Glane23, Gobonobo, Hu12, IceCreamAntisocial, Isnow, JFY, Jaimetex, Jared, Jeyradan, John254, KC Panchal, Katieh5584, Kazkaskazkasako, Kehrbkyid, Khauser32353, Lantonov, Lexor, Little Mountain 5, Luci Sandor, MER-C, Mariabrenna, Melamed katz, MichaelJanich, Michaelas10, Mushroom, Mykon, NighthawkJ, Nina Gerlach, Noneforall, Nono64, Obscurans, PDH, Proquence, Raz1el, Rosieredfield, RunOrDie, RyanGerbil10, Sabedon, Seb951, Steinsky, StrangeTaco, Strych, TedE, Teo64x, TestPilot, Tokek, Vary, Versus22, Vgcap, Vojtech.dostal, Wisdom89, WriterHound, Wwfboy17, Xmort, Yamamoto Ichiro, Ychastnik APL, Yersinia, Zav, 179 anonymous edits

Reverse transcriptase *Source:* <http://en.wikipedia.org/w/index.php?oldid=352994896> *Contributors:* Abuttlmao, Affl23a, Agathman, AhmadH, Alethiareg, Alnokta, Arcadian, Boghog2, CarinaT, Conscious, Conversion script, David Shear, Dkabban-GMU, Dr. Strangelove, Drphilharmonic, Dullhunk, Dwmyers, Echinoidea, Espresso Addict, Eteklema-GMU, Eumeng.chong, Evenprime, Flower-gmu, Fratrep, Fvasconcellos, GATHrawn23, Gaurav, Graft, Hariva, Iknowyourider, JWSchmidt, Jfdwolff, Jrdioko, Kevin Forsyth, Km2452-GMU, Lancevortex, Lantonov, M1ssIontomas2k4, Magnus Manske, Master son, Mikael Häggström, Morpheios Melas, Myanw, Mystaker1, Naj-GMU, PDH, PFHLai, Paul Drye, Peak, Pearle, Peter Znamenskiy, Physicstjedi, Rebroad, RedWolf, RelentlessRecusant, Richard Arthur Norton (1958-), Richwil, Rodasmith, Serephine, Skhanal-GMU, Suisui, Svick, Thomasgl, Tjmayerinsf, Wejstheman, Wimt, WojPob, WriterHound, 97 anonymous edits

DNA microarray *Source:* <http://en.wikipedia.org/w/index.php?oldid=354703792> *Contributors:* 168..., 198.143.250.xxx, 3 Löwi, Aciel, AdamRetchless, Adenosine, Afluegel, AlastairM3754, Albinpaulxavier, Amar kamath, Amarilla858, Amosolarin, Andrew73, Andrewericoleman, Angr, Apers0n, Arcadian, Aremith, Artgen, Aurelduv, AxelBoldt, Barticus88, Bender235, Bensaccount, Big Bird, Bioinformin, Blacksun, Brian Crawford, Brooke618, Calmargulis, Carlsbad, Cathleenmrocco, Ceyockey, Chemist234, Ciar, Cmungall, Cobalt137cc, Conversion script, Crissmyass, DGG, Dacharle, Davidweiss, Deniz13, Dicklyon, Discospinster, DoctorDNA, Dr. William Jacobs, Drdaveng, Duncharris, Effeitsanders, Elagatis, Environmatt, Fawcett5, Figma, GV wiki, Geelo, Geno-Supremo, Giftlite, GiollaUidir, Glane23, Glashedy, GoldenTorc, Graham87, Gurmukh.s, Gustavocarra, HankMansion, Hephastos, Hgfernan, Hraefen, Hu12, Hydkat, Iidnormal, Ilovecmicroarrays, Ipodamos, Jack-A-Roe, Jackhynes, Jafield, Jahiegel, Jeff Marks, Jethero, Jfdwolff, Jgreene1305, Jgruszynski, JonHarder, Jonathan Hall, Jondel, JosephBarillari, Jotomicron, Jubal, Kantokano, Karin sandby, Khalid hassani, KirbyRandolf, Klamber, Kurykh, La goutte de pluie, Lantonov, Larssono, Lassefolkersen, Lea Cleary, Lexor, Lfrench, Lightmouse, Lindsay658, Lisa230579, Lord.lucan, Luwo, Malljaja, Marj Tiefert, MatthewBChambers, Mattpope, Michael Hardy, Miguel Andrade, Movado73, Mxn, Natarajangesan, NicolasStransky, Nieselt, NoQuarter, Nrhoads, Nuptsey, Olympos, Orlyal, PCAexplorer, Paphrag, Parijata, Parkinson, Patrick Maitland, Peipei, PeterCanthropus, Pgan002, Phix7, Pixelface, Postglock, Pvosta, Qinatan, Quartertone, Radagast83, RainbowOfLight, Raul654, Rebut, Rich Farmbrough, Rlockner, Ruud no1, Sandraesugi, Schutz, Scott.spillman, Sentausa, Shabd sound, Shamrocktuesday, Skittleys, Slustbader, Speedyboy, Spongebobsqpants, Squidoni, Srlasky, SteveChervitzTrutane, Suffusion of Yellow, Tameeria, Terrace4, TestPilot, Thdog42, The Sunshine Man, TheNhibition, Thingg, Thomas.Hentrich, Thomas81, Thripsi, Thumperward, TimShell, Toddmartinsky, Tombadog, Toniosky, Tstokes, Tstrobaugh, Tuskaloosa, Twisp, Typochimp, Tysi, Unint, Vegetarianrage, Venullian, Vladimir Pilný, Vrr, WAS 4.250, WMod-NS, Wavelength, Webridge, West Brom 4ever, Whosasking, Willia, Wk muriithi, Wleizero, Wli625, Woohookitty, WriterHound, Yaki-gaijin, Zoolium, Zven, 356 anonymous edits

Triple-stranded DNA *Source:* <http://en.wikipedia.org/w/index.php?oldid=344800290> *Contributors:* Jacopo Werther, Kaarel, PDH, Rubicon, Rumping, TAKE, Tomgally, InfoCan, 14 anonymous edits

G-quadruplex *Source:* <http://en.wikipedia.org/w/index.php?oldid=354225125> *Contributors:* Acer, Bcbpts, Clicketyclack, Darklilac, Derek.cashman, Harold f, Hynca-Hooley, Iridos, Jlh29, Jmrivort, Loki, Luuva, Nono64, Pbgaga, PigFlu Oink, Pimprenelle, Rich Farmbrough, Seans Potato Business, Simon12, Snori, Talon Artaine, TimVickers, Vinodscaria, Ynnib, 101 anonymous edits

DNA Analysis *Source:* <http://en.wikipedia.org/w/index.php?oldid=355247815> *Contributors:* Abdulrahman1998, AdRock, Addshore, Aetkin, Ale jrb, Alexius08, Amorymeltzer, Angelliaaaa, Angielaj, Antalas, Apers0n, Arakunem, ArchonMagnus, Arjun01, Artgen, Arvindn, Bassbonerocks, Bayerischermann, Beatgoddess, Beland, Biotechmaster, BloodGrapefruit2, Bobo192, BorisTM, Bradjamesbrown, Canadian-Bacon, Celefin, Ceyockey, Curedh, D, DBaba, Daniel 1992, Daniellakasteel, Davewild, Dawn Bard, Dbfirs, Debresser, Delldot, Dentong93, DoctorDNA, Docu, Doodlefrog, Doyley, EastTN, Electric goat, Epr123, Esparkhu, Flewis, Forluvoft, Fsostrain09, Funny.pochi, Gail, Giftlite, Gilliam, GlassCobra, Gogo Dodo, Goodnightmush, Grealat386, H.sanat, HaLoGuY007, Hardyplants, Hede2000, Homer Landskirty, II MusLiM HyBRiD II, IW.HG, IceCreamAntisocial, InvictaHOG, loeth, J.delanoy, J04n, JaGa, JackPercival, Jamiesongj, Jasynnash2, Jayden54, Jdrewitt, Jfdwolff, John Hill, Johnuniq, JonHarder, Kamielvdz, Kevin143, Kjkolb, KramarDanlaku, Lambourgreeny, Languorous Lass, Lawrence Cohen, Leevanjackson, Lethaniol, Lexor, Little Mountain 5, Llort, Luckas Blade, Luk, MONGO, Martian.knight, Marysunshine, Melisa River, Merlion444, Mmr173, Mrdallaway, Mschel, NewEnglandYankee, Nicradcliffe, Nihiltres, Novangelis, Osarius, Pearle, Pelago, PoeticVerse, RainbowOfLight, Recognizance, Reconsider the static, Remember, Rjwilmsi, Rmhermen, Rsabbatini, SEIBasaurus, Sam Hayes, Sander123, Sarah.arora, Sarah00001, Sciencechick, Senator Palpatine, Sepeople, Sferrier, Sferry, Shenme, Shimgray, Sijupadhyay, Sjö, SoWhy, Spinctrl, Spitfire, Spook`, Stepp-Wulf, Superm401, The Anome, The Thing That Should Not Be, The blue inferno, Tide rolls, Uncle Dick, Utcursch, Venu62, Versus22, Wdfarmer, Werdan7, Wimt, Winchelsea, Woohookitty, Wouterstomp, WriterHound, XApple, Z22, Zodon, 392 anonymous edits

Human genome *Source:* <http://en.wikipedia.org/w/index.php?oldid=354578138> *Contributors:* A-giau, AFriedman, Acather96, Aitias, Alan Au, Alansohn, Andux, Aniketripathi, Anner w88, Apers0n, Arch dude, Aukicha, AxelBoldt, BMF81, Beland, Benbest, Bkkbrad, Blanchardb, Bogey97, BravozuluSA, Btunell, Burn, Ceyockey, Chrispounds, Christopherlin, Clamster5, ClockworkSoul, Cptmurdok, Dan100, Darklilac, David D., DavidRader, Daycd, Dbabbitt, Dbachmann, Dbrouse, Deconstructhis, DerHexer, Dfrg.msc, Discospinster, Diza, Dougofborg, Dymorodrepanis, Ec5618, Efbfweborg, Elb2000, Eleassar, Eliz81, Elockid, Epr123, Esparkhu, EthicsGradient, Evil saltine, Evolver, Excirial, Fenice, Filnik, Foregan, Forluvoft, Frankenpuppy, Fred Hsu, FreplySpang, Fullmetal Alchemist Fan, GT5162, Genetics411, Gh5046, GideonKlok, Giftlite, Gilliam, Giraffedata, Graft, Greensburger, Grendelkhan, GreyWyvern, Grylliade, Gwicky, Gwrightnyc, Haseo9999, Icey, Imnotminkus, InternetMeme, J.delanoy, JWSchmidt, Jessicanr, JoanneB, Joeshmodo, John Kenney, Johnraciti, Jrockley, Juled, Karenjc, Keegscee, Kocio, Kosigrim, Kozuch, Kranthikiranakula, Leevanjackson, Little Mountain 5, LittleHow, LittleOldMe, Llamallama, Lost.goblin, MER-C, MONGO, Malleus Fatuorum, MarcK, Marek69, Markus Kuhn, Matt.T, Maurice Carbonaro, Mccready, Michael Hardy, Michael Snow, Miguel Andrade, Mikael Häggström, Mike Lin, Moorens, MrBird, Mushin, Mxn, MyNameIsNotBob, Mystery rubberduck, NBeale, NHRHS2010, NawlinWiki, Nbauman, Nectarflowed, Nemesis of Reason, Nicholas Cimini, Nihiltres, Nosce, Nothingofwater, Oblivious, Opabinia regalis, Oudguy, Oxymoron83, P b1999, PDH, Paughsw, Pharaoh of the Wizards, Planetscape, Plasticup, Pleasantville, Plindenbaum, Plociam, Ppgardne, Prestonmag, Quizkajer, RFerreira, RJASE1, RJHall, Rabidphage, Rakela, Rdsmith4, Rich Farmbrough, Rjwilmsi, RobHutten, Samsara, Sasha I, Saturation2, Sayeth, Severa, Shanes, Shwot, Snowmanradio, Spencer, SriMesh, Srose, Stanisla0, Avogadro, Svetovid, Syrup17, TJHawk5789, TedE, Testmann, The Anome, The Eagle, The Obento Musubi, TheXenocide, TimVickers, Tpradbury, Tryangle of Hope, Turnstep, Ugur Basak, VMS Mosaic, Versus22, Viddin4, Vigeesh, Violetriga, Vlado4, Vojtech.dostal, Vsb, Vuvur1, Vviralpapillome, Wet dog fur, Whiner01, Wlodzimierz, Wouterstomp, Wpf pokefan, Wsiegmund, Xkeops, Ypetrachenko, Zafiroblue05, Zero g, ^demon, 327 anonymous edits

Image Sources, Licenses and Contributors

File:ADN animation.gif *Source:* http://en.wikipedia.org/w/index.php?title=File:ADN_animation.gif *License:* Public Domain *Contributors:* Aushulz, Bawolff, Brian0918, Elecbullet, Kersti Nebelsiek, Magadan, Mattes, Origamiemensch, Stevenfruitsmaak, Str4nd, Túrelío, 5 anonymous edits

File:Loudspeaker.svg *Source:* <http://en.wikipedia.org/w/index.php?title=File:Loudspeaker.svg> *License:* Public Domain *Contributors:* Bayo, Gmaxwell, Husky, Iamunknown, Nethac DIU, Omegatron, Rocket000, 5 anonymous edits

Image:DNA chemical structure.svg *Source:* http://en.wikipedia.org/w/index.php?title=File:DNA_chemical_structure.svg *License:* Creative Commons Attribution-Sharealike 2.5 *Contributors:* Madprime, Wickey, 1 anonymous edits

File:DNA orbit animated static thumb.png *Source:* http://en.wikipedia.org/w/index.php?title=File:DNA_orbit_animated_static_thumb.png *License:* GNU Free Documentation License *Contributors:* 84user adapting file originally uploaded by Richard Wheeler (Zephyris) at en.wikipedia

Image:GC DNA base pair.svg *Source:* http://en.wikipedia.org/w/index.php?title=File:GC_DNA_base_pair.svg *License:* Public Domain *Contributors:* User:Isilanes

Image:AT DNA base pair.svg *Source:* http://en.wikipedia.org/w/index.php?title=File:AT_DNA_base_pair.svg *License:* Public Domain *Contributors:* User:Isilanes

Image:A-DNA, B-DNA and Z-DNA.png *Source:* http://en.wikipedia.org/w/index.php?title=File:A-DNA,_B-DNA_and-Z-DNA.png *License:* GNU Free Documentation License *Contributors:* Original uploader was Richard Wheeler (Zephyris) at en.wikipedia

Image:Parallel telomere quadruple.png *Source:* http://en.wikipedia.org/w/index.php?title=File:Parallel_telomere_quadruple.png *License:* GNU Free Documentation License *Contributors:* User:Splette

Image:Branch-dna.png *Source:* <http://en.wikipedia.org/w/index.php?title=File:Branch-dna.png> *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* Original uploader was Peter K. at en.wikipedia

Image:Multi-branch-dna.png *Source:* <http://en.wikipedia.org/w/index.php?title=File:Multi-branch-dna.png> *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* . Original uploader was Peter K. at en.wikipedia

Image:Cytosine chemical structure.png *Source:* http://en.wikipedia.org/w/index.php?title=File:Cytosine_chemical_structure.png *License:* GNU Free Documentation License *Contributors:* BorisTM, Bryan Derksen, Cacycle, Cwbn (commons), Edgar181, Engineer gena, 1 anonymous edits

Image:5-methylcytosine.png *Source:* <http://en.wikipedia.org/w/index.php?title=File:5-methylcytosine.png> *License:* GNU Free Documentation License *Contributors:* EDUCA33E, Luigi Chiesa, Mysid

Image:Thymine chemical structure.png *Source:* http://en.wikipedia.org/w/index.php?title=File:Thymine_chemical_structure.png *License:* GNU Free Documentation License *Contributors:* Arrowsmaster, BorisTM, Bryan Derksen, Cacycle, Edgar181, Leyo

Image:Benzopyrene DNA adduct 1JDG.png *Source:* http://en.wikipedia.org/w/index.php?title=File:Benzopyrene_DNA_adduct_1JDG.png *License:* GNU Free Documentation License *Contributors:* Benjah-bmm27, Bstlee, 1 anonymous edits

Image:T7 RNA polymerase at work.png *Source:* http://en.wikipedia.org/w/index.php?title=File:T7_RNA_polymerase_at_work.png *License:* GNU Free Documentation License *Contributors:* User:Splette

Image:DNA replication en.svg *Source:* http://en.wikipedia.org/w/index.php?title=File:DNA_replication_en.svg *License:* Public Domain *Contributors:* user:LadyofHats

Image:Nucleosome 2.jpg *Source:* http://en.wikipedia.org/w/index.php?title=File:Nucleosome_2.jpg *License:* Public Domain *Contributors:* Original uploader was TimVickers at en.wikipedia

Image:Nucleosome (opposites attracts).JPG *Source:* [http://en.wikipedia.org/w/index.php?title=File:Nucleosome_\(opposites_attracts\).JPG](http://en.wikipedia.org/w/index.php?title=File:Nucleosome_(opposites_attracts).JPG) *License:* Public Domain *Contributors:* Illustration by David S. Goodsell of The Scripps Research Institute (see this site)

Image:Lambda repressor 1LMB.png *Source:* http://en.wikipedia.org/w/index.php?title=File:Lambda_repressor_1LMB.png *License:* GNU Free Documentation License *Contributors:* Original uploader was Zephyris at en.wikipedia

Image:EcoRV 1RVA.png *Source:* http://en.wikipedia.org/w/index.php?title=File:EcoRV_1RVA.png *License:* GNU Free Documentation License *Contributors:* Original uploader was Zephyris at en.wikipedia

Image:Holliday Junction cropped.png *Source:* http://en.wikipedia.org/w/index.php?title=File:Holliday_Junction_cropped.png *License:* GNU Free Documentation License *Contributors:* Original uploader was TimVickers at en.wikipedia

Image:Holliday junction coloured.png *Source:* http://en.wikipedia.org/w/index.php?title=File:Holliday_junction_coloured.png *License:* GNU Free Documentation License *Contributors:* Original uploader was Zephyris at en.wikipedia

Image:Chromosomal Recombination.svg *Source:* http://en.wikipedia.org/w/index.php?title=File:Chromosomal_Recombination.svg *License:* Creative Commons Attribution 2.5 *Contributors:* User:Gringer

Image:DNA nanostructures.png *Source:* http://en.wikipedia.org/w/index.php?title=File:DNA_nanostructures.png *License:* unknown *Contributors:* (Images were kindly provided by Thomas H. LaBean and Hao Yan.)

File:Francis Crick.png *Source:* http://en.wikipedia.org/w/index.php?title=File:Francis_Crick.png *License:* unknown *Contributors:* Photo: Marc Lieberman

File:Rosalind Franklin.jpg *Source:* http://en.wikipedia.org/w/index.php?title=File:Rosalind_Franklin.jpg *License:* unknown *Contributors:* Avraham, Feydey, Hannes Röst, Shabd sound

File:Raymond Gosling.jpg *Source:* http://en.wikipedia.org/w/index.php?title=File:Raymond_Gosling.jpg *License:* unknown *Contributors:* Original uploader was Davidruben at en.wikipedia

Image:DNA_sequence.png *Source:* http://en.wikipedia.org/w/index.php?title=File:DNA_sequence.png *License:* Public Domain *Contributors:* LadyofHats, Llull, Sjef

Image:Mutation Surveyor Trace.jpg *Source:* http://en.wikipedia.org/w/index.php?title=File:Mutation_Surveyor_Trace.jpg *License:* Public Domain *Contributors:* User:BotMultichill

Image:Sequencing.jpg *Source:* <http://en.wikipedia.org/w/index.php?title=File:Sequencing.jpg> *License:* GNU Free Documentation License *Contributors:* John Schmidt

Image:DNA Sequencin 3 labeling methods.jpg *Source:* http://en.wikipedia.org/w/index.php?title=File:DNA_Sequencin_3_labeling_methods.jpg *License:* Public Domain *Contributors:* Original uploader was Abizar at en.wikipedia

Image:Radioactive Fluorescent Seq.jpg *Source:* http://en.wikipedia.org/w/index.php?title=File:Radioactive_Fluorescent_Seq.jpg *License:* Creative Commons Attribution-Sharealike 2.5 *Contributors:* Original uploader was Abizar at en.wikipedia

Image:CE Basic.jpg *Source:* http://en.wikipedia.org/w/index.php?title=File:CE_Basic.jpg *License:* Public Domain *Contributors:* Original uploader was Abizar at en.wikipedia

Image:Sanger sequencing read display.gif *Source:* http://en.wikipedia.org/w/index.php?title=File:Sanger_sequencing_read_display.gif *License:* Public Domain *Contributors:* Loris

Image:DNA Sequencing gDNA libraries.jpg *Source:* http://en.wikipedia.org/w/index.php?title=File:DNA_Sequencing_gDNA_libraries.jpg *License:* Public Domain *Contributors:* Abizar

Image:D1S80Demo.gif *Source:* <http://en.wikipedia.org/w/index.php?title=File:D1S80Demo.gif> *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* PaleWhaleGail, uploaded to Wikimedia by Charvosi

Image:DNA polymerase.png *Source:* http://en.wikipedia.org/w/index.php?title=File:DNA_polymerase.png *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* User:Yikrazuul

Image:DNA polymerase.svg *Source:* http://en.wikipedia.org/w/index.php?title=File:DNA_polymerase.svg *License:* Creative Commons Attribution-Sharealike 2.5 *Contributors:* User:Madprime

Image:Topoisomerase I.png *Source:* http://en.wikipedia.org/w/index.php?title=File:Topoisomerase_I.png *License:* Public Domain *Contributors:* Illustration by David S. Goodsell of The Scripps Research Institute (see this site)

Image:DNA-(A)80-model.png *Source:* [http://en.wikipedia.org/w/index.php?title=File:DNA-\(A\)80-model.png](http://en.wikipedia.org/w/index.php?title=File:DNA-(A)80-model.png) *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* User:P99am

Image:FirstSketchOfDNADoubleHelix.jpg *Source:* <http://en.wikipedia.org/w/index.php?title=File:FirstSketchOfDNADoubleHelix.jpg> *License:* unknown *Contributors:* Carcharoth, Lion Wilson, Mxn, PDH, Solipsist, Twas Now, 5 anonymous edits

Image:DNA-structure-and-bases.png *Source:* <http://en.wikipedia.org/w/index.php?title=File:DNA-structure-and-bases.png> *License:* Public Domain *Contributors:* AutisticPsycho2, MesserWoland, Origamiemensch

Image:Dna-split.png *Source:* <http://en.wikipedia.org/w/index.php?title=File:Dna-split.png> *License:* Public Domain *Contributors:* Conscious, Dietzel65, Juliancolton, LadyofHats, Madprime, Magnus Manske, Niki K, Pixeltoo, Roybb95, Samulili, Teetaweepo, Thomas81, 2 anonymous edits

Image:Spinning DNA.gif *Source:* http://en.wikipedia.org/w/index.php?title=File:Spinning_DNA.gif *License:* Public Domain *Contributors:* USDA

Image:Dna-split2.png *Source:* <http://en.wikipedia.org/w/index.php?title=File:Dna-split2.png> *License:* Public Domain *Contributors:* Bobarino, Denniss, Samulili, Wobble

Image:Sarfus.DNABiochip.jpg *Source:* <http://en.wikipedia.org/w/index.php?title=File:Sarfus.DNABiochip.jpg> *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* Nanolane

File:DNA chemical structure.svg *Source:* http://en.wikipedia.org/w/index.php?title=File:DNA_chemical_structure.svg *License:* Creative Commons Attribution-Sharealike 2.5 *Contributors:* Madprime, Wickey, 1 anonymous edits

File:DNA-fragment-3D-vdW.png *Source:* <http://en.wikipedia.org/w/index.php?title=File:DNA-fragment-3D-vdW.png> *License:* Public Domain *Contributors:* Benjah-bmm27

File:Simple harmonic oscillator.gif *Source:* http://en.wikipedia.org/w/index.php?title=File:Simple_harmonic_oscillator.gif *License:* Public Domain *Contributors:* User:Oleg Alexandrov

File:3D model hydrogen bonds in water.jpg *Source:* http://en.wikipedia.org/w/index.php?title=File:3D_model_hydrogen_bonds_in_water.jpg *License:* GNU Free Documentation License *Contributors:* User:snek01

File:ABDNAXrgpj.jpg *Source:* <http://en.wikipedia.org/w/index.php?title=File:ABDNAXrgpj.jpg> *License:* GNU Free Documentation License *Contributors:* I.C. Baianu et al.

File:Four-way_DNA_junction.gif *Source:* http://en.wikipedia.org/w/index.php?title=File:Four-way_DNA_junction.gif *License:* Public Domain *Contributors:* Aushulz, Molatwork, Origamiensch, TimVickers, 1 anonymous edits

File:Parallel telomere quadruple.png *Source:* http://en.wikipedia.org/w/index.php?title=File:Parallel_telomere_quadruple.png *License:* GNU Free Documentation License *Contributors:* User:Splette

File:Telomer-structure.gif *Source:* <http://en.wikipedia.org/w/index.php?title=File:Telomer-structure.gif> *License:* GNU Free Documentation License *Contributors:* Original uploader was Iridos at en.wikipedia

File:G-quadruplex gene regulation.jpg *Source:* http://en.wikipedia.org/w/index.php?title=File:G-quadruplex_gene_regulation.jpg *License:* Creative Commons Attribution-Sharealike 2.5 *Contributors:* Julian Huppert

File:DNA in water.jpg *Source:* http://en.wikipedia.org/w/index.php?title=File:DNA_in_water.jpg *License:* GNU Free Documentation License *Contributors:* User:Bbkkk

File:X ray diffraction.png *Source:* http://en.wikipedia.org/w/index.php?title=File:X_ray_diffraction.png *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* Thomas Spletstoesser

File:Bragg diffraction.png *Source:* http://en.wikipedia.org/w/index.php?title=File:Bragg_diffraction.png *License:* GNU General Public License *Contributors:* user:hadmack

File:X Ray Diffractometer.JPG *Source:* http://en.wikipedia.org/w/index.php?title=File:X_Ray_Diffractometer.JPG *License:* GNU Free Documentation License *Contributors:* FF02::3, Pieter Kuiper

File:Dna-SNP.svg *Source:* <http://en.wikipedia.org/w/index.php?title=File:Dna-SNP.svg> *License:* Creative Commons Attribution 2.5 *Contributors:* User:Gringer

File:DNA_replication_en.svg *Source:* http://en.wikipedia.org/w/index.php?title=File:DNA_replication_en.svg *License:* Public Domain *Contributors:* user:LadyofHats

File:DNA Model Crick-Watson.jpg *Source:* http://en.wikipedia.org/w/index.php?title=File:DNA_Model_Crick-Watson.jpg *License:* Public Domain *Contributors:* User:Alkivar

File:Museo Principe Felipe. ADN.jpg *Source:* http://en.wikipedia.org/w/index.php?title=File:Museo_Principe_Felipe_ADN.jpg *License:* Creative Commons Attribution-Sharealike 2.0 *Contributors:* Fernando

File:DNA Under electron microscope Image 3576B-PH.jpg *Source:* http://en.wikipedia.org/w/index.php?title=File:DNA_Under_electron_microscope_Image_3576B-PH.jpg *License:* unknown *Contributors:* Original uploader was SeanMack at en.wikipedia

File:A-B-Z-DNA Side View.png *Source:* http://en.wikipedia.org/w/index.php?title=File:A-B-Z-DNA_Side_View.png *License:* Public Domain *Contributors:* Original uploader was Thorwald at en.wikipedia

File:BU Bio5.jpg *Source:* http://en.wikipedia.org/w/index.php?title=File:BU_Bio5.jpg *License:* Creative Commons Attribution-Sharealike 2.0 *Contributors:* Original uploader was Elapied at fr.wikipedia

File:SLAC detector edit1.jpg *Source:* http://en.wikipedia.org/w/index.php?title=File:SLAC_detector_edit1.jpg *License:* GNU Free Documentation License *Contributors:* User:Mfield, User:Starwiz

File:ISIS exptal hall.jpg *Source:* http://en.wikipedia.org/w/index.php?title=File:ISIS_exptal_hall.jpg *License:* unknown *Contributors:* wurzeller

File:Silica.svg *Source:* <http://en.wikipedia.org/w/index.php?title=File:Silica.svg> *License:* Public Domain *Contributors:* Silica.jpg:

File:Genomics GTL Pictorial Program.jpg *Source:* http://en.wikipedia.org/w/index.php?title=File:Genomics_GTL_Pictorial_Program.jpg *License:* Public Domain *Contributors:* Mdd

File:RNA pol.jpg *Source:* http://en.wikipedia.org/w/index.php?title=File:RNA_pol.jpg *License:* Public Domain *Contributors:* InfoCan

File:Primase 3B39.png *Source:* http://en.wikipedia.org/w/index.php?title=File:Primase_3B39.png *License:* Public Domain *Contributors:* own work

File:MGMT+DNA 1T38.png *Source:* http://en.wikipedia.org/w/index.php?title=File:MGMT+DNA_1T38.png *License:* Public Domain *Contributors:* own work

File:DNA damaged by carcinogenic 2-aminofluorene AF .jpg *Source:* http://en.wikipedia.org/w/index.php?title=File:DNA_damaged_by_carcinogenic_2-aminofluorene_AF.jpg *License:* Public Domain *Contributors:* Brian E. Hingerty, Oak Ridge National Laboratory Suse Broydre, New York University Dinshaw J. Patel, Memorial Sloan Kettering Cancer Center

File:DNA Repair.jpg *Source:* http://en.wikipedia.org/w/index.php?title=File:DNA_Repair.jpg *License:* Public Domain *Contributors:* Courtesy of Tom Ellenberger, Washington University School of Medicine in St. Louis.

File:A-DNA orbit animated small.gif *Source:* http://en.wikipedia.org/w/index.php?title=File:A-DNA_orbit_animated_small.gif *License:* GNU Free Documentation License *Contributors:* User:Bstlee, User:Zephyris

File:Plasmid emNL.jpg *Source:* http://en.wikipedia.org/w/index.php?title=File:Plasmid_emNL.jpg *License:* GNU Free Documentation License *Contributors:* Denniss, Glenn, Rasbak

File:Circular DNA Supercoiling.png *Source:* http://en.wikipedia.org/w/index.php?title=File:Circular_DNA_Supercoiling.png *License:* GNU Free Documentation License *Contributors:* Richard Wheeler (Zephyris)

File:Chromatin chromosom.png *Source:* http://en.wikipedia.org/w/index.php?title=File:Chromatin_chromosom.png *License:* Public Domain *Contributors:* User:Magnus Manske

File:Chromosome.svg *Source:* <http://en.wikipedia.org/w/index.php?title=File:Chromosome.svg> *License:* GNU Free Documentation License *Contributors:* User:Dietzel65, User:Magnus Manske, User:Tryphon

File:Chr2 orang human.jpg *Source:* http://en.wikipedia.org/w/index.php?title=File:Chr2_orang_human.jpg *License:* Creative Commons Attribution-Sharealike 2.5 *Contributors:* Verena Schubel, Stefan Müller, Department Biologie der Ludwig-Maximilians-Universität München.

File:3D-SIM-3 Prophase 3 color.jpg *Source:* http://en.wikipedia.org/w/index.php?title=File:3D-SIM-3_Prophase_3_color.jpg *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* Lothar Schermelleh

File:Chromosome2 merge.png *Source:* http://en.wikipedia.org/w/index.php?title=File:Chromosome2_merge.png *License:* Public Domain *Contributors:* Original uploader was Evercat at en.wikipedia

File:Transkription Translation 01.jpg *Source:* http://en.wikipedia.org/w/index.php?title=File:Transkription_Translation_01.jpg *License:* Public Domain *Contributors:* User:Kuebi

File:RibosomaleTranskriptionsEinheit.jpg *Source:* <http://en.wikipedia.org/w/index.php?title=File:RibosomaleTranskriptionsEinheit.jpg> *License:* GNU Free Documentation License *Contributors:* User:Merops

File:Chromosome Conformation Capture Technology.jpg *Source:* http://en.wikipedia.org/w/index.php?title=File:Chromosome_Conformation_Capture_Technology.jpg *License:* Public Domain *Contributors:* User:Kangyun1985

File:Mitochondrial DNA and diseases.png *Source:* http://en.wikipedia.org/w/index.php?title=File:Mitochondrial_DNA_and_diseases.png *License:* Creative Commons Attribution-Sharealike 2.5 *Contributors:* User:XXXL1986

File:PCR.svg *Source:* <http://en.wikipedia.org/w/index.php?title=File:PCR.svg> *License:* Creative Commons Attribution-Sharealike 2.5 *Contributors:* User:Madprime

File:Pcr gel.png *Source:* http://en.wikipedia.org/w/index.php?title=File:Pcr_gel.png *License:* GNU Free Documentation License *Contributors:* Habj, Ies, PatríciaR, Retama, Saperaud

File:Dividing Cell Fluorescence.jpg *Source:* http://en.wikipedia.org/w/index.php?title=File:Dividing_Cell_Fluorescence.jpg *License:* unknown *Contributors:* Will-moore-dundee

File:SFP discovery principle.jpg *Source:* http://en.wikipedia.org/w/index.php?title=File:SFP_discovery_principle.jpg *License:* Free Art License *Contributors:* User:Agbiotec

File:Cdnaarray.jpg *Source:* <http://en.wikipedia.org/w/index.php?title=File:Cdnaarray.jpg> *License:* Creative Commons Attribution 2.5 *Contributors:* Mangapoco

File:Expression of Human Wild-Type and P239S Mutant Palladin.png *Source:* http://en.wikipedia.org/w/index.php?title=File:Expression_of_Human_Wild-Type_and_P239S_Mutant_Palladin.png *License:* unknown *Contributors:* see above

File:Co-dominance Rhododendron.jpg *Source:* http://en.wikipedia.org/w/index.php?title=File:Co-dominance_Rhododendron.jpg *License:* Creative Commons Attribution 2.0 *Contributors:* Ayacop, Bestiasonica, Cillas, FlickrLickr, FlickreviewR, Horcha, Kanonkas, Kevmin, MPF, Para

File:Bloodcell_sun_flares_pathology.jpeg *Source:* http://en.wikipedia.org/w/index.php?title=File:Bloodcell_sun_flares_pathology.jpeg *License:* Public Domain *Contributors:* Birindand, Karelj, NEON ja, 1 anonymous edits

File:Sarfus.DNABiochip.jpg *Source:* <http://en.wikipedia.org/w/index.php?title=File:Sarfus.DNABiochip.jpg> *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* Nanolane

File:DNA nanostructures.png *Source:* http://en.wikipedia.org/w/index.php?title=File:DNA_nanostructures.png *License:* unknown *Contributors:* (Images were kindly provided by Thomas H. LaBean and Hao Yan.)

Image:DNA replication split.svg *Source:* http://en.wikipedia.org/w/index.php?title=File:DNA_replication_split.svg *License:* Creative Commons Attribution-Sharealike 2.5 *Contributors:* User:Madprime

Image:1axc tricolor.png *Source:* http://en.wikipedia.org/w/index.php?title=File:1axc_tricolor.png *License:* GNU Free Documentation License *Contributors:* Opabinia regalis

Image:Cell Cycle 2.png *Source:* http://en.wikipedia.org/w/index.php?title=File:Cell_Cycle_2.png *License:* GNU Free Documentation License *Contributors:* Original uploader was Zephyris at en.wikipedia

Image:brokechromo.jpg *Source:* <http://en.wikipedia.org/w/index.php?title=File:Brokechromo.jpg> *License:* GNU Free Documentation License *Contributors:* Multichill, Square87, Wickey, 2 anonymous edits

Image:Ssvsds.jpg *Source:* <http://en.wikipedia.org/w/index.php?title=File:Ssvsds.jpg> *License:* GNU Free Documentation License *Contributors:* Bkell, Square87

Image:Uracil base glycosidase.jpg *Source:* http://en.wikipedia.org/w/index.php?title=File:Uracil_base_glycosidase.jpg *License:* Public Domain *Contributors:* Original uploader was TimVickers at en.wikipedia

Image:DNA Repair.jpg *Source:* http://en.wikipedia.org/w/index.php?title=File:DNA_Repair.jpg *License:* Public Domain *Contributors:* Courtesy of Tom Ellenberger, Washington University School of Medicine in St. Louis.

Image:Dnarepair1.png *Source:* <http://en.wikipedia.org/w/index.php?title=File:Dnarepair1.png> *License:* unknown *Contributors:* Harold Brenner

Image:Dnadamage.png *Source:* <http://en.wikipedia.org/w/index.php?title=File:Dnadamage.png> *License:* unknown *Contributors:* Harold Brenner

Image:Molbio-Header.svg *Source:* <http://en.wikipedia.org/w/index.php?title=File:Molbio-Header.svg> *License:* Public Domain *Contributors:* User:Squidonius

Image:Ribosome mRNA translation en.svg *Source:* http://en.wikipedia.org/w/index.php?title=File:Ribosome_mRNA_translation_en.svg *License:* Public Domain *Contributors:* User:LadyofHats

File:simple transcription initiation1.svg *Source:* http://en.wikipedia.org/w/index.php?title=File:Simple_transcription_initiation1.svg *License:* Public Domain *Contributors:* User:Forluvoft

File:simple transcription elongation1.svg *Source:* http://en.wikipedia.org/w/index.php?title=File:Simple_transcription_elongation1.svg *License:* Public Domain *Contributors:* User:Forluvoft

File:simple transcription termination1.svg *Source:* http://en.wikipedia.org/w/index.php?title=File:Simple_transcription_termination1.svg *License:* Public Domain *Contributors:* User:Forluvoft

File:RetroTranscription.jpg *Source:* <http://en.wikipedia.org/w/index.php?title=File:RetroTranscription.jpg> *License:* Public Domain *Contributors:* Cyberugo

Image:Reverse_Transcriptase_1HMV.png *Source:* http://en.wikipedia.org/w/index.php?title=File:Reverse_Transcriptase_1HMV.png *License:* Public Domain *Contributors:* User:Boghog2

Image:Reverse transcription.svg *Source:* http://en.wikipedia.org/w/index.php?title=File:Reverse_transcription.svg *License:* Creative Commons Attribution 3.0 *Contributors:* Filip em

Image:Zidovudine.svg *Source:* <http://en.wikipedia.org/w/index.php?title=File:Zidovudine.svg> *License:* Public Domain *Contributors:* User:Fvasconcellos

Image:Microarray2.gif *Source:* <http://en.wikipedia.org/w/index.php?title=File:Microarray2.gif> *License:* Public Domain *Contributors:* Original uploader was Paphrag at en.wikipedia

File:NA_hybrid.svg *Source:* http://en.wikipedia.org/w/index.php?title=File:NA_hybrid.svg *License:* Public Domain *Contributors:* User:Squidonius

File:Microarray exp horizontal.svg *Source:* http://en.wikipedia.org/w/index.php?title=File:Microarray_exp_horizontal.svg *License:* Public Domain *Contributors:* User:Squidonius

Image:Affymetrix-microarray.jpg *Source:* <http://en.wikipedia.org/w/index.php?title=File:Affymetrix-microarray.jpg> *License:* Creative Commons Attribution 2.5 *Contributors:* Schutz

Image:Microarray-schema.jpg *Source:* <http://en.wikipedia.org/w/index.php?title=File:Microarray-schema.jpg> *License:* Public Domain *Contributors:* Larssono, Million Moments

Image:Heatmap.png *Source:* <http://en.wikipedia.org/w/index.php?title=File:Heatmap.png> *License:* Public Domain *Contributors:* Original uploader was Miguel Andrade at en.wikipedia

Image:G-quadruplex.gif *Source:* <http://en.wikipedia.org/w/index.php?title=File:G-quadruplex.gif> *License:* Public Domain *Contributors:* Original uploader was Harold f at en.wikipedia

Image:Telomer-structure.gif *Source:* <http://en.wikipedia.org/w/index.php?title=File:Telomer-structure.gif> *License:* GNU Free Documentation License *Contributors:* Original uploader was Iridos at en.wikipedia

Image:G-quadruplex gene regulation.jpg *Source:* http://en.wikipedia.org/w/index.php?title=File:G-quadruplex_gene_regulation.jpg *License:* Creative Commons Attribution-Sharealike 2.5 *Contributors:* Julian Huppert

Image:Karyotype.png *Source:* <http://en.wikipedia.org/w/index.php?title=File:Karyotype.png> *License:* Public Domain *Contributors:* Courtesy: National Human Genome Research Institute

License

Creative Commons Attribution-Share Alike 3.0 Unported
<http://creativecommons.org/licenses/by-sa/3.0/>
